

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Genome wide recruitment of TREX reveals a direct interaction with the Rpb1 CTD



Dominik Matthias Meinel
aus
Straubing, Deutschland

2013

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Frau Dr. Katja Sträßer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 25.10.2013

Meinel Dominik

Dissertation eingereicht am:	25.10.2013
Erstgutachterin:	Dr. Katja Sträßer
Zweitgutachter:	Prof. Dr. Patrick Cramer
Mündliche Prüfung am:	9.12.2013

Summary

Gene expression is a complex and highly regulated process, in which mRNA synthesis and export are tightly linked, to ensure a streamlined process of gene expression. The molecular mechanisms of the coupling of transcription to mRNP formation and export are not fully understood until to date. It is known, that in *S. cerevisiae* the conserved TREX complex couples transcription to mRNA export and mediates mRNP formation, but how TREX recruitment is orchestrated with transcription and if all proteins involved in mRNA export are recruited to all genes is still largely unknown.

In this study, all key players which recruit the mRNA export receptor Mex67-Mtr2 to the mRNP were analyzed for their recruitment on a genome-wide level by ChIP-chip. All subunits of the TREX complex are recruited to all RNAPII genes. Moreover, the two alternative mRNA export receptor adaptors Npl3 and Nab2 are recruited to all RNAPII genes. Additionally, Nab2 is recruited to RNAPIII genes, indicating a function in RNAPIII metabolism. The THO subunits, Sub2 and Yra1 exhibit a length dependent recruitment, *i.e.* their occupancy increases from 5' to the 3' of the gene. A novel ribozyme-based assay showed that the TREX complex as well as Nab2 and Npl3 are dependent on the nascent mRNA in their recruitment, but that the length dependency of TREX is not mediated by RNA. Analysis of several general elongation factors indicated, that the length dependency is unique to TREX, but that the increase is in accordance with Rpb1-CTD tyrosine 1 and serine 2 phosphorylation pattern. In fact, correct phosphorylation of the CTD residues serine 2 and tyrosine 1 are needed *in vivo* for efficient TREX recruitment as shown by ChIP analysis of Rpb1-CTD mutants. Interestingly, both phosphorylations seem to be interdependent *in vivo*. Importantly, pull down experiments showed, that the THO subcomplex of TREX directly binds to serine 2 and serine 5 diphosphorylated repeats *in vitro* possibly explaining length dependent recruitment. This also demonstrates that Yra1, which was recently shown to bind to the Rpb1-CTD, is not needed to recruit the THO complex to the CTD. Furthermore, by disrupting the CTD interaction domain of Yra1, THO/TREX recruitment to the genes is not influenced *in vivo* and the Yra1 truncation mutant does not associate anymore with the TREX complex, indicating, that THO/TREX is recruited to the transcription site by direct CTD interaction of THO with the serine 2 and serine 5 di-phosphorylated CTD. Transcriptome analysis of a TREX recruitment mutant, which abolishes length dependent recruitment of the whole TREX complex, but does not influence RNAPII levels on the genes, revealed for the first time that a gradual increase of a protein complex is important for correct expression of the genome. Furthermore, we found that the CTR of the general elongation factor Spt5 might be involved in mRNP formation and export. We found that the Spt5-CTR affects Tho1 and THO/TREX recruitment and mutations of the CTR lead to decreased Tho1 but increased THO/TREX levels at the genes. Overexpression of Tho1 negatively regulates THO recruitment, indicating a regulatory role of Tho1 in THO/TREX recruitment, one of the key steps in early mRNP biogenesis.

Taken together, we elucidated the molecular recruitment mechanism of TREX to the transcription site by the CTD of the RNAPII. We propose that TREX functions to keep the mRNA in the vicinity of the transcription machinery by interacting with the RNAPII CTD as well as with the nascent mRNA. This could serve to ensure correct processing and mRNP formation.

Publications

During this thesis I contributed to the following publications:

Meinel DM, Burkert-Kautzsch C, Kieser A, O'Duibhir E, Siebert M, Mayer A, Cramer P, Söding J, Holstege FCP, Sträßer K (2013) Recruitment of TREX to the Transcription Machinery by its Direct Binding to the Phospho-CTD of RNA Polymerase II, PLOS Genetics, 9 (11)

Schenk L, Meinel DM, Sträßer K and Gerber AP (2012) La-motif-dependent mRNA association with Slf1 promotes copper detoxification in yeast. RNA 18 (3)

Chanarat S, Burkert-Kautzsch C, Meinel DM and Sträßer K (2012) Prp19C and TREX: interacting to promote transcription elongation and mRNA export. Transcription 3 (1)

Content

I. Introduction	6
1. Gene expression	6
2. The Transcription cycle	6
3. mRNP formation	7
4. The THO/TREX complex	9
5. Nab2 and Npl3 – alternative mRNA export receptor adaptors	12
6. The THSC/TREX-2 complex is involved in mRNA export	13
7. Tho1 is involved in mRNA export	14
8. Coupling of transcription, mRNP formation and mRNA export	14
8.1 Orchestrating co-transcriptional events by the Rpb1-CTD	15
8.2 The RNA as recruitment platform	17
8.3 Spt5 - an upcoming recruitment platform	18
8.4 Additional ways to couple steps in mRNP biogenesis	18
9. mRNA export	19
10. Aims and scope of this study	20
II. Results	22
1. TREX recruitment	22
2. TREX, Nab2 and Npl3 recruitment to noncoding RNAs	26
3. THO /TREX is recruited length dependent	29
4. TREX recruitment is RNA dependent	34
5. CTD phosphorylation is essential for TREX recruitment <i>in vivo</i>	36
6. THO is recruited to the genes independently of Yra1	40
7. The THO complex binds directly to Ser2-Ser5 diphosphorylated CTD repeats	43
8. C-terminal TAP-tag fusion to Tho2 impairs TREX recruitment	45
9. The 5' to 3' increasing recruitment of THO/TREX is important for correct gene expression	48
10. The PCID of Yra1 genetically interacts with <i>THO2-TAP</i>	48
11. Tho1 recruitment recapitulates THO/TREX recruitment	49
12. The Spt5 CTR regulates recruitment of mRNA export factors	52
13. Tho1 overexpression decreases THO/TREX recruitment	54
III. Discussion	57
1. The mRNA export receptor adapters TREX, Nab2 and Npl3 are recruited to all genes	57
2. Recruitment of the TREX complex	58
3. Is the recruitment of TREX to the phospho-CTD of Rpb1 conserved?	62
4. Possible implications of increasing TREX occupancy for mRNP biogenesis	63
5. Gbp2 and Hrb1 may be recruited by THO to the CTD and then transferred to the mRNA	64
6. The Spt5-CTR: a possible role in mRNP biogenesis and export	65

7.	Conclusions.....	66
IV.	Material and Methods	68
1.	Materials.....	68
1.1	Consumables and Chemicals	68
1.2	Equipment	68
1.3	Yeast strains.....	69
1.4	Plasmids.....	71
1.5	Oligo-nucleotide sequences	72
1.6	peptides	76
1.7	Buffers, solutions and Growth media.....	76
1.8	Antibodies.....	77
2.	Methods	79
2.1	General techniques	79
2.1.1	Molecular cloning.....	79
2.1.2	PCR.....	79
2.1.3	Special cloning strategies	80
2.1.3.1	Cloning of the ribozyme constructs.....	80
2.1.3.2	Cloning of Y1F CTD mutants	80
2.1.3.3	Cloning of Yra1- Δ PCID	81
2.1.4	SDS PAGE	81
2.1.5	Western-Blot	81
2.2	Yeast culture	82
2.2.1	Culture of <i>Saccharomyces cerevisiae</i>	82
2.2.2	Transformation of yeast cells	82
2.2.3	Whole cell extracts (WCE)	82
2.3	TAP-Purification.....	83
2.4	CTD-Peptide Pull-down Assay	84
2.5	Small scale ChIP	84
2.6	qPCR.....	85
2.7	ChIP-chip.....	86
2.7.1	ChIP.....	86
2.7.3	Array hybridization	87
2.8	Data analysis of ChIP-chip Data.....	88
4.8.2	Gene occupancy profiles	88
2.8.3	Meta profiles of gene classes	88
2.8.4	Length dependency plots	89
2.8.5	Correlations of recruitment.....	89
2.9	Expression profiling of Tho2-TAP	90

Content

Abbreviations	91
Acknowledgements	92
Tables	93
Figures	94
References.....	96

I. Introduction

1. Gene expression

Gene expression is a fundamental process in every living cell that translates the information stored in the DNA into proteins. Eukaryotic cells contain a nucleus, where the genetic information is stored and transcribed by RNA Polymerase II (RNAPII) into messenger RNAs. These need to be exported to the cytoplasm for the translation into protein. Regulation of the gene expression process ensures the flexibility of the cell to adapt with a given situation.

2. The Transcription cycle

Gene expression has its starting point with the transcription of the gene. Transcription itself begins with the tightly regulated process of initiation. RNAPII together with the help of a multitude of transcription factors synthesizes the pre-mRNA. During the transcription cycle, which consists of the main phases initiation, elongation and termination, many processes take place and the phases have fluent transitions. First, the initiation factors TFIIA, TFIIB, TFIID, TFII E, TFII F and TFII H are recruited to the gene promoter and form together with RNAPII the preinitiation complex (PIC) (Hahn 2004). These processes are highly regulated for flexible gene expression, *e.g.* by the mediator complex, a 1.4 MDa multiprotein complex in yeast (Bourbon 2008), which is a co-activator transmitting signals from distantly bound upstream activators to the core promoter (Vojnic et al. 2011).

After the DNA duplex is opened, transcription by RNAPII can initiate and the RNAPII starts to synthesize the RNA (Hahn 2004). At this stage of transcription, the RNA synthesis works error prone and up to the first 8 nucleotides several abortive rounds of initiation might take place, until the early elongation phase is reached (Margaritis and Holstege 2008). During elongation phase many *bona fide* transcription elongation factors support the RNAPII in productively synthesizing the nascent mRNA. One obstacle that the transcription machinery has to master is that in the eukaryotic cell the DNA is not “free” but packaged into chromatin. This can be visualized in electron micrographs as “beads on a string”, the so called 10nm fiber, which consists of the DNA wrapped around the nucleosomes (Olins et al. 2003, Rando and Ahmad 2007). To be able to transcribe through this barrier many elongation factors help to remodel the chromatin. Among these transcription factors are Spt4/5, the Paf1 complex and Elf1. Another important elongation factor is TFII S, which enables RNAPII to read through blocks of elongation, by stimulating the RNAPII for backtracking and cleaving of the faulty mRNA piece to reactivate RNAPII for transcription (reviewed in Cramer 2004).

At the polyA site one more transition of the transcription cycle takes place: The nascent pre-mRNA is released by cleavage of the RNA. RNAPII continues to transcribe downstream of the polyA site until the RNAPII leaves the chromatin by the termination process. For termination two models were suggested, one called “allosteric model”, which is based on changes of the elongation complex downstream of the cleavage event that lead to loss of positive elongation factors. Secondly, the “torpedo model”, in which the remainders of the RNA after the polyA site cleavage are degraded by Rat1, which catches up to RNAPII and thereby releases, together with other factors like Rai1 and Rtt103, RNAPII from the chromatin (for review see Richard and Manley 2009, Proudfoot 2011) .

3. mRNP formation

The newly synthesized RNA is processed and packed into a messenger ribonucleoprotein particle (mRNP) by several events that create a mature mRNP ready for export from the nucleus. To ensure that all processes can take place in a timely correct manner, the first steps in mRNP formation start already co-transcriptionally (mRNP formation is summarized in Figure 1).

One of the first hall marks of mRNP formation is the capping of the 5' end of the nascent RNA. The m⁷-G-cap is made by first removing the terminal phospho-group from the nascent transcript, and adding a GTP under the loss of a pyrophosphate resulting in a 5'-5' triphosphate linkage (reviewed in Topisirovic et al. 2011). Later on, the N7 of the GTP is methylated and the cap is bound by the cap binding complex (Cbp80 and Cbp20 in yeast). The cap is important to protect the transcript from exonucleolytic degradation and promotes several downstream processes of gene expression: splicing in higher eukaryotes, mRNA export and even translation via the cytoplasmic cap binding protein eIF4E (reviewed in Topisirovic et al. 2011).

The second important step in mRNP formation is the removal of introns of the primary transcript by splicing. Splicing is a complex process, which involves several protein-RNA complexes. These catalyze in a highly regulated way the trans-esterification process, which removes the introns and ligates the two adjacent exons together (reviewed in Will and Luhrmann 2011). Interestingly, introns are very rare in yeast genes. Interestingly, although only *app.* 4% of the yeast genes contain one or more introns, *app.* 27% of the yeast mRNAs originate from intron containing genes, reflecting their high expression levels (Ares, JR et al. 1999). It is though that the removal of introns serves as a quality checkpoint for correct mRNA formation in metazoans, while incorrect splicing leads to nonsense-mediated mRNA decay (reviewed in Chang et al. 2007).

In parallel, many RNA binding proteins are recruited to the mRNA to form the mature mRNPs. Those proteins have various functions, like regulation of other processes in mRNP formation, possible roles in mRNA export, in translation, in protecting the mRNP from degradation, in proper mRNP packing or in protection of hybridization of the nascent mRNA back to the DNA and thereby causing hyper-recombination (reviewed in Rondón et al. 2010).

Subsequently, the mRNA is released from the transcription site by cleavage of the pre-mRNA by the cleavage and 3' end processing complex, which consists in yeast of the Cleavage and Polyadenylation Factor (CPF) and the Cleavage Factors I A and B (CFIA and CFIB). This complex process involves binding of these factors to the upstream of the cleavage site located efficiency element (EE) and positioning element (PE) in the nascent mRNA (reviewed in Mandel et al. 2008). The cleavage site itself is generally defined by a sequence element containing a pyrimidine followed by an adenosine stretch (Zhao et al. 1999). Cleavage is coupled to poly-adenylation *in vivo* by harboring both enzymatic activities in the 3' end processing complex, to ensure a high efficiency. The 3' end is poly-adenylated by the polyA-polymerase (Pap1), which needs the interaction with other members of the 3' end processing complex for defining the proper length of the poly(A) tail (Mandel et al. 2008). Upon poly-adenylation, polyA binding proteins, *e.g.* Pab1 and Nab2, are recruited to the polyA tail, which typically consists in yeast of around 70-90 adenosines and in human of 200-300 adenosines (Mandel et al. 2008). The length of the polyA tail is also known to influence mRNA stability and thereby the half-life of the mRNA. The polyA binding protein 1 (Pab1) has a function in controlling the poly-adenylation reaction, for a proper length of the polyA tail. Furthermore the polyA RNA binding protein Nab2 seems to also have a function in polyA length control (Hector et al. 2002).

Interestingly, most genes not only have one cleavage and polyA site, but several (Wilkening et al. 2013, Ozsolak et al. 2010). Especially in higher eukaryotes polyA site selection can influence gene expression by in- or excluding miRNA binding sites to the 3' region of a gene and thereby fine tuning expression levels (Sandberg et al. 2008). Shortening and thereby loss of such regulatory sites might even activate oncogenes (Mayr and Bartel 2009). Alternative cleavage and polyadenylation might also lead to truncated proteins and to omitted binding sites for RNA binding proteins, which in turn lead to different transcript stability or even to a different localization (see Tian and Manley 2013 for review).

After the assembled mRNP is released from the site of transcription and its 3' end is processed, the mRNP is ready for export. The mRNA export takes place when the export receptor Mex67-Mtr2 is recruited to the mRNP.

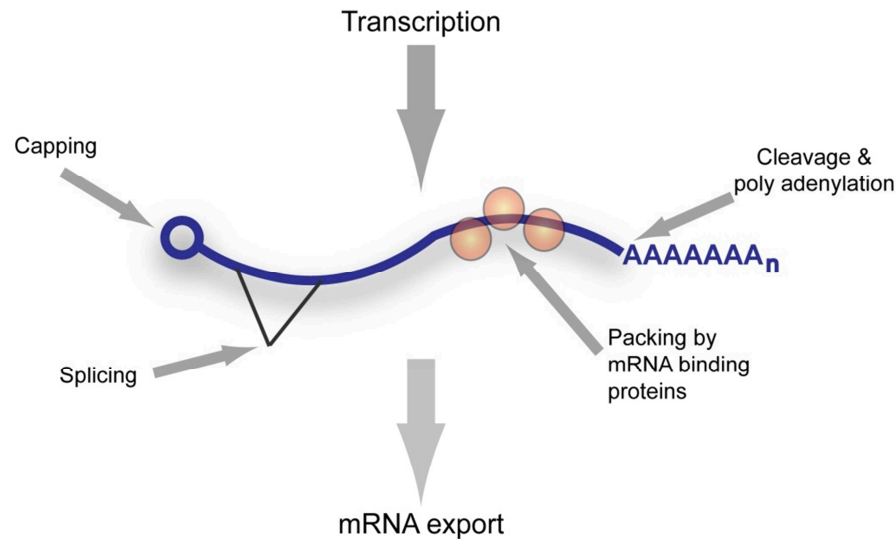


Figure 1: early mRNP formation

mRNP formation has its starting point during transcription with the nascent RNA (blue). The capping, splicing, packaging by mRNA binding proteins and the cleavage and polyadenylation process contribute to the formation of a mature mRNP in the nucleus, which is competent for mRNA export into the cytoplasm. All processes are highly coupled, and are mostly co-transcriptionally to ensure a highly efficient gene expression.

4. The THO/TREX complex

mRNA export begins in yeast already during transcription with the co-transcriptional recruitment of the TREX complex to the active chromatin. The TREX complex, which couples transcription and export, is highly conserved throughout eukaryotes and consists in yeast of the THO subcomplex and the proteins Sub2, Yra1, Gbp2 and Hrb1. THO is a hetero-pentameric subcomplex and consist of Tho2, Hpr1, Mft1, Thp2 and Tex1 (Strässer et al. 2002, Chávez et al. 2000). THO was initially found to promote transcription elongation and to protect the cell from transcription dependent hyper-recombination in yeast (Chávez et al. 2000, Chavez et al. 2001) as well as in higher eukaryotes (Domínguez-Sánchez et al. 2011). The function of THO/TREX is thereby highly connected to mRNP formation, since the hyper-recombination seems to be based on the formation of R-loops. Those are formed by nascent mRNA molecules, which fold back to the DNA and form together a stable hybrid, which can cause replication impairment or single strand breaks in the single DNA strand which is not hybridized to the RNA. This seems to take place when no proper mRNP is formed due to the lack of TREX, which prevents the back folding by packing the mRNA into an mRNP (Huertas and Aguilera 2003, Domínguez-Sánchez et al. 2011). Recently, an EM structure of the THO complex showed the elongated architecture of the THO complex and the same study also showed the complex's ability to bind to ssDNA, dsDNA and RNA *in vitro* (Peña et al. 2012).

In addition to the THO subcomplex, the TREX complex comprises the two SR proteins Gbp2, Hrb1, the DECD-box Helicase Sub2 and Yra1. The latter two form a stable dimer, which binds to the mRNP (Strässer and Hurt 2000, Strässer and Hurt 2001 and Strässer et al. 2002). Although THO subunit knock out in yeast is not lethal, but impairs growth and at higher temperatures mRNA export, Yra1 and Sub2 are both essential for the cell. Studies with temperature sensitive mutants of *YRA1* and *SUB2* revealed a similar or even worse hyper-recombination phenotype as for THO subunit knock out (Jimeno et al. 2002).

The TREX complex is thought to transfer its RNA binding subunits, Gbp2, Hrb1, Yra1 and Sub2 to the mRNP after co-transcriptional recruitment to the active chromatin (Strässer et al. 2002, Abruzzi et al. 2004). Recently Dbp2, a DEAD box helicase, was shown to be important to prepare the RNA for the transfer of Yra1 by unwinding secondary structures (Ma et al. 2013). Reed and Hurt (2002) proposed that after the early mRNP is formed, a remodeling event takes place, in which the DECD box helicase Sub2 leaves the mRNP and Yra1 directly interacts with the conserved mRNA export receptor Mex67-Mtr2. Thus, Yra1 serves as an export receptor adaptor (Strässer and Hurt 2000). The interaction of Mex67 with Yra1 is also conserved in higher eukaryotes. Yra1 is subsequently ubiquitinated by Tom1 at the nuclear face of the NPC, leaves the mRNP and the mRNP traverses through the central pore of the NPC (Iglesias et al. 2010).

Interestingly, it was published, that not only the TREX component Yra1, but also Hpr1, a subunit of the THO complex, might recruit Mex67-Mtr2 to the mRNP via its ubiquitylation. Hpr1 can be ubiquitinated by Rsp5 and then interact with the C-terminal ubiquitin associated domain of Mex67. However, at the moment this recruitment mechanism for Mex67 was only shown for a subset of inducible genes (Gal2, Gal10 and Hsp104) and whether this is a general mechanism or a bypass is not known yet (Hobeika et al. 2009, Hobeika et al. 2007, Gwizdek et al. 2006).

In higher eukaryotes, it is thought that the TREX complex is not recruited by the transcription machinery as in yeast, but during splicing by the spliceosome (Masuda et al. 2005). This happens most likely also co-transcriptionally and human THO/TREX subunits were found to copurify with the spliceosome (Zhou et al. 2002, Chen et al. 2007). Furthermore, human TREX can interact via its subunit Aly (Yra1) with the Cap binding complex protein CBP80 and is thereby recruited to the 5' end of the transcript (Cheng et al. 2006). Nevertheless, recent data showed that TREX is also recruited non-splicing dependent (Nojima et al. 2007, Katahira et al. 2009, Lei et al. 2013), opening the possibility for a similar recruitment of TREX by the transcription machinery in human as in yeast.

Recently, in yeast a direct link of Yra1 with the cleavage and polyadenylation machinery was discovered by physical interaction of Yra1 with Pcf11 (Johnson et al. 2009). This is in line with a faulty

3' end processing in THO/Sub2 mutants, which were shown to cause formation of dense chromatin (Rougemaille et al. 2008) and display inefficient poly-adenylation (Saguez et al. 2008). Curiously, Yra1 and Pcf11 interaction was shown to be independent of Sub2/TREX (Johnson et al. 2009), suggesting possible alternative Yra1 recruitment pathways or an independent function.

Although much is known about the function of the THO/TREX complex, we still do not know how TREX is recruited to the transcription site on a molecular basis. It was recently shown, that Yra1 might interact with the Rpb1-CTD (MacKellar and Greenleaf 2011). But considering that Yra1 was shown to interact via the same domain with the Rpb1-CTD, with Sub2 (THO/TREX) and also with Pcf11 (Johnson et al. 2009, Johnson et al. 2011) and that mutations in this domain lead to mis-localization of Yra1 to the cytoplasm (MacKellar and Greenleaf 2011), it is at the moment far from clear, whether the CTD interaction of Yra1 recruits the TREX complex at all or serves as an alternative mechanism to recruit only Yra1. Furthermore, it was recently shown, that the Prp19 complex, a protein complex involved in splicing, is needed to stabilize THO/TREX occupancy *in vivo* at the 3' end of the gene, but this interaction does not explain the recruitment mechanism of TREX (Chanarat et al. 2011). Thus how TREX is recruited on a molecular basis is a big, remaining question.

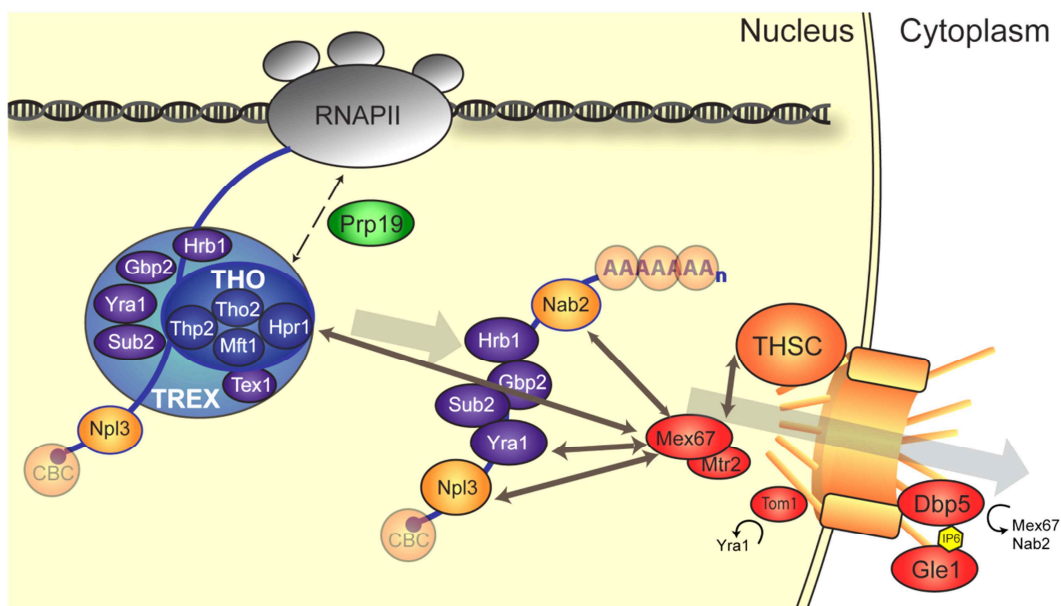


Figure 2: The yeast mRNA export pathway.

THO/TREX in blue is recruited co-transcriptionally. It is stabilized at the 3' end of the gene by the Prp19 complex. The TREX subunits Sub2, Yra1, Hrb1 and Gbp2 are known to associate with the mRNP. Hpr1 and Yra1 are able to directly recruit the export-receptor Mex67-Mtr2 (red) to the mRNP and thereby mediate mRNA export. Mex67-Mtr2 is also recruited via Nab2 and Npl3 (orange), two co-transcriptionally associating proteins. Nab2 interacts therefore also with Yra1. The THSC complex (orange), which is located at the NPC (orange) can also attract Mex67-Mtr2 to the mRNP. At the nuclear face of the NPC, Tom1 dissociates Yra1 from the mRNP by ubiquitinylation of Yra1. After the mRNP traversed to the cytoplasm, the DEAD box Helicase Dbp5, which associates already inside the nucleus to the mRNP, but is also recruited at the cytoplasmic face of the NPC, dissociates Mex67-Mtr2 and Nab2 from the mRNP at the cytoplasmic face of the NPC. The DEAD-box helicase Dbp5 mediates remodeling of the mRNP and Dbp5 is activated by IP₆ (yellow) and Gle1 (red) in an ATP dependent manner on the cytoplasmic face of the NPC and releases thereby the mRNA into the cytoplasm for translation. The mRNA export factors are subsequently re-imported into the nucleus by importins (Karyopherins) for a new cycle of mRNA export.

5. Nab2 and Npl3 – alternative mRNA export receptor adaptors

Additional to Yra1 and Hpr1, there are two other proteins implicated in recruiting Mex67-Mtr2 to the mRNP. One of those is Nab2, whose overexpression was shown to repress the lethal *Δyra1* phenotype (Vinciguerra et al. 2005). Nab2 was initially found to bind to poly-adenylated RNA and to be involved in the length control of the polyA tail (Hector et al. 2002). Nab2 is highly conserved in higher eukaryotes (ZC3H14 in human) and harbors an RGG repeat domain and seven tandem CCCH Zn-finger motifs, which were shown to bind to polyA-RNA (Anderson et al. 1993, Kelly et al. 2007 and Kuhlmann et al. 2013). Nab2 also serves as an mRNA export receptor adaptor, by direct interaction with Mex67. Interestingly, this interaction is stimulated by Yra1, which dissociates from the ternary Nab2-Mex67-Yra1 complex after Yra1 di-ubiquitylation by Tom1 at the nucleoplasmic face of the NPC (Iglesias et al. 2010). It was also shown that a direct interaction of Nab2 with Mlp1, a NPC protein, is needed for the export of the mRNA and serves as one of the first contacts of the mRNP to the NPC (Fasken et al. 2008). Subsequently, Nab2 shuttles together with the mRNP to the cytoplasm (Iglesias et al. 2010). At the cytoplasmic face of the NPC, the DEAD-box helicase Dbp5 remodels ATP and Gle1 dependent the mRNP and dissociates thereby Nab2 (Tran et al. 2007). The Karyopherin Kap104 binds in turn to Nab2 and mediates Ran-GTP dependent reimported into the nucleus for a new round of mRNA export (Lee and Aitchison 1999). Interestingly, Pab1, a polyA binding protein with similar function in control of polyA-tail length, stays associated with the mRNP until translation and functions in translation initiation (Kessler and Sachs 1998).

Npl3 is another protein involved in mRNA export by recruiting Mex67-Mtr2 to the mRNP for mRNA export. Additional to its function in mRNP export, Npl3 was recently shown to be involved in the export of the large ribosomal subunit via its direct interaction with the 5S rRNA, similar to Mex67 (Yao et al. 2007), but interestingly without a need for Mex67 in Npl3 mediated export of the large ribosomal subunit (Hackmann et al. 2011). Npl3 is a SR like protein, which is very similar to the TREX proteins Gbp2 and Hrb1 in domain architecture. SR proteins are defined by containing N-terminal RNA recognition motifs (RRMs) and a C-terminus containing RS rich regions which are reversibly phosphorylated (Graveley and Hertel). All three proteins were shown to shuttle to the cytoplasm and stay associated with the mRNP during translation (Windgassen and Krebber 2003, Häcker and Krebber 2004, Windgassen et al. 2004). Interestingly, Npl3 architecture shares homology to proteins of the SR and hnRNP families: hnRNP proteins usually carry di-methylations of closely repeated RGG tripeptides at their C-terminus additional to their N-terminal RRM (Liu and Dreyfuss 1995). Npl3 carries N-terminally two RRM and C-terminally six RGG tripeptides, which were shown to be methylated and two SR repeats, which can be phosphorylated (Gilbert et al. 2001, McBride et al. 2005, Yun and Fu 2000). As Gbp2 and Hrb1, Npl3 was also shown to be co-transcriptionally recruited

to active chromatin and to bind to the mRNA (Lei et al. 2001). It is thought that Npl3 is recruited to the transcription site by binding to the phosphorylated CTD of Rpb1 via its C-terminal part (Dermoddy et al. 2008), while Gbp2 and Hrb1 are recruited via the TREX complex (Häcker and Krebber 2004, Hurt et al. 2004). Furthermore, Npl3 was shown to promote transcription elongation by antagonizing 3' end formation in yeast, especially by competing with Rna15 for binding sites on the nascent RNA (Bucheli and Buratowski 2005, Deka et al. 2008). Furthermore, Npl3 was shown to be involved in splicing, by promoting recruitment of the spliceosome (Kress et al. 2008). During transcription elongation Npl3 is phosphorylated on Ser 411 by CK2 and thus handed over to the mRNP, since CTD binding affinity of Npl3 is reduced by the phosphorylation (Dermoddy et al. 2008). Additionally, binding to Rna15 recognition sites on the RNA is reduced while general, sequence unspecific RNA binding of Npl3 is not affected by the phosphorylation (Dermoddy et al. 2008). Thus Npl3 does not compete with Rna15 anymore and cleavage and polyadenylation of the mRNA can take place. Upon Npl3 dephosphorylation on residue serine 411 by Glc7, Npl3 and Mex67-Mtr2 can directly interact for mRNA export. The timing of this dephosphorylation event is still unknown, but given the interaction of Glc7 with the Cleavage and Polyadenylation factor (CPF), it seems likely, that 3' end processing triggers the dephosphorylation of Npl3 and thereby stimulates the interaction with Mex67-Mtr2 (Gilbert and Guthrie 2004b). The export receptor is thereby tethered to the mRNP and mRNA export can take place after cleavage and polyadenylation (Gilbert et al. 2001, Gilbert and Guthrie 2004a). Npl3 shuttles together with the mRNP from the nucleus to the cytoplasm, where it also carries out a function in translation by negatively regulating translation elongation (Windgassen et al. 2004). Subsequently, Npl3 is phosphorylated in the cytoplasm by Sky1, dissociates from the mRNP, and is then imported back to the nucleus by the importin Mtr10 (Senger et al. 1998, Gilbert et al. 2001).

6. The THSC/TREX-2 complex is involved in mRNA export

TREX-2 or THSC is another protein complex involved in mRNP biogenesis and mRNA export. It is located at the nucleoplasmic face of the NPC and consists of Thp1, Sac3, Cdc31, Sus1 and Sem1 (Faza et al. 2009, reviewed in Köhler and Hurt 2007). THSC mutants exhibit, similar to THO/TREX mutants, a transcription dependent hyper recombination and transcription phenotype (Gallardo and Aguilera 2001, Gallardo 2003). The THSC complex might be linked to transcription via sharing its subunit Sus1 with the transcriptionally active SAGA complex (Rodríguez-Navarro et al. 2004) Though sharing several similarities, THSC and TREX differ in several aspects: THSC is located at the nuclear pore and TREX is not (Lei et al. 2003), Sub2 overexpression rescues the THO mutations, but is synthetic lethal

with THSC mutants (Jimeno et al. 2002, Gonzalez-Aguilera et al. 2008) and Nab2 overexpression rescues the mRNA export defect of Thp1 (THSC) deletion but does not rescue THO/TREX mutants (Jimeno et al. 2006, Gallardo 2003). THSC was also implicated in gene-gating, a process in which highly induced genes are re-located for a more efficient mRNA export to the nuclear pore (Cabal et al. 2006, Rodríguez-Navarro and Hurt 2011).

7. Tho1 is involved in mRNA export

The hnRNP Tho1 was initially identified as a high copy suppressor of the transcription and export phenotype of THO complex but not THSC complex mutants in yeast (Jimeno et al. 2006). Although Tho1 does not copurify in yeast with THO/TREX, it was recently shown, that the human orthologous protein CIP29 copurifies with TREX in an ATP dependent manner (Dufu et al. 2010). CIP29 was shown to bind to UAP56 (Sub2) and its very close paralog DDX39 in a two-hybrid assay (Leaw et al. 2004). Although it was shown that CIP29 is able to regulate the helicase activity of DDX39 and most likely also of UAP56, the function of CIP29 is not yet clarified (Sugiura et al. 2007). Yeast Tho1 is recruited co-transcriptionally to active chromatin, in an RNA dependent manner (Jimeno et al. 2006). The NMR structure of the SAP (scaffold-associated protein) domain in the C-terminal half of Tho1 was recently published (Dodson et al. 2010) and it was shown to be needed for suppression of Hpr1 deletion by Tho1 overexpression (THO/TREX) and to bind to double stranded DNA *in vitro* (Jimeno et al. 2006). The C-terminal half of Tho1, which contains no known domains, binds strongly to RNA *in vitro*. Furthermore, Tho1 deletion rescues the phenotype of the *nab2-1* mutant (Jimeno et al. 2006). However, at the moment the exact function of Tho1 in mRNA export is not known.

8. Coupling of transcription, mRNP formation and mRNA export

To ensure that the complex and multifaceted processes of forming mature mRNPs and delivering them to the cytoplasm are efficient and not error-prone, all processes taking place from transcription to export are coupled and highly interconnected. This starts already during early transcription elongation: as soon as the nascent mRNA 5' end appears out of the core RNAPII, capping takes place. The now following steps are packing of the mRNA into the mRNP, co-transcriptional splicing and 3'end formation. But how does the cell orchestrate these diverse processes? For this the cell uses three different recruitment platforms at the level of transcription: The RNAPII CTD, the RNA and the Spt5-CTR.

8.1 Orchestrating co-transcriptional events by the Rpb1-CTD

Rpb1, the largest subunit of RNAPII, has an extended C-terminal domain (CTD) which is thought to be largely unstructured. The CTD consists of heptad-repeats, with a consensus sequence of YSPTSPS. The CTD has 26 repeats in yeast and 52 in human (for review see Zhang et al. 2012). Not all repeats are strictly conserved and cells with mutated repeats are still viable, but not all repeats are equally important for the cell (West and Corden 1995).

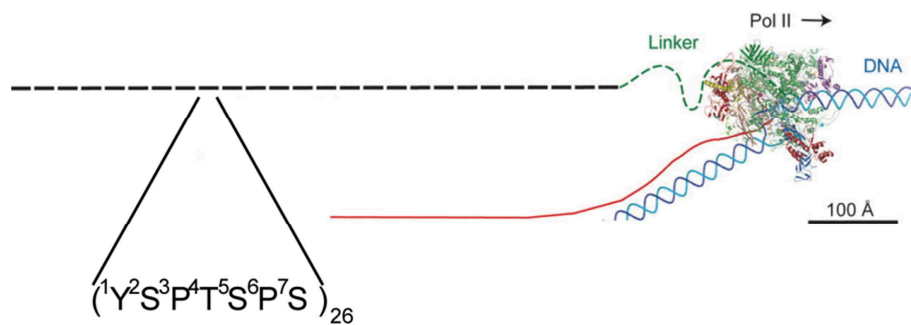


Figure 3: The Rpb1-CTD

The largest subunit of the RNAPII, Rpb1, has an extended CTD which consist in yeast of 26 and in human of 52 heptad repeats. The consensus sequence is Y-S-P-T-S-P-S. The RNAPII structure is given as ribbon diagram, the DNA is depicted in blue, the nascent RNA in red and the extended CTD is shown in black. Figure adapted from (Meinhart et al. 2005).

The Rpb1-CTD repeats contain several residues which are phosphorylated and differential phosphorylation turns the CTD into a versatile recruitment platform: The CTD repeats are phosphorylated on Ser5 during transcription initiation. Phosphorylation is carried out mainly by Kin28 but also by Srb10 in yeast, in higher eukaryotes by Cdk7 and Cdk8, respectively (Zhang et al. 2012). The Ser5 phosphorylation is known to be involved in coordinating early chromatin remodeling events (*e.g.* by recruiting Set1) and also to set the starting point for co-transcriptional mRNP formation: The capping enzyme complex subunit Ceg1 (in yeast), which serves as guanylyltransferase and possibly the methyltransferase (Abd1 in yeast) can directly interact with the Ser5 phosphorylated CTD (Cho et al. 1997, for review see Zhang et al. 2012). Although the capping enzyme in higher eukaryotes is structurally different, it also interacts with the Ser5 phosphorylated CTD for their recruitment (Fabrega et al. 2003, Ghosh et al. 2011). Moreover, Ser7 of the CTD repeats was shown to be phosphorylated in yeast and higher eukaryotes during transcription initiation and Kin28 was shown to be a Ser7 kinase (for review see Zhang et al. 2012). It is known so far, that Ser7 phosphorylation is needed for transcription and correct processing of human snRNA genes (Egloff et al. 2012), but the exact role of Ser7 phosphorylation is still under investigation.

After promoter clearance, the phosphorylation pattern of the Rpb1-CTD changes to recruit proteins which act during transcription elongation. This transition is marked by Ser2 phosphorylation of the CTD repeats. In higher eukaryotes this is highly regulated by the Negative Elongation Factor (NELF), which causes promoter proximal pausing and DSIF, which reactivates the stalled polymerase. In yeast no promoter proximal pausing is known, and immediately after promoter clearance the Bur1 kinase (in higher eukaryotes Cdk9) is recruited by binding to phosphorylated Ser5 and initial Ser2 phosphorylation is carried out (Qiu et al. 2009). Additionally, Bur1 was recently shown to phosphorylate Ser7 (Tietjen et al. 2010, Bataille et al. 2012). After priming the CTD by initial Ser2 phosphorylation for recruitment of the Ctk1 kinase complex (Cdk12), which is the major Ser2 kinase, the CTD repeats are extensively phosphorylated on the Ser2 residues (Cho et al. 2001). Furthermore, Tyr1 was recently shown to be massively phosphorylated by an unknown kinase during elongation (Mayer et al. 2012a). Simultaneously, Ser5 phosphorylation marks are removed by the yeast phosphatase Rtr1 and Ssu72, but a basal Ser5 phosphorylation persists throughout the open reading frame (Krishnamurthy et al. 2004, Mosley et al. 2009, Kim et al. 2009, Bataille et al. 2012). The Ser2 phosphorylation is well known to orchestrate many processes of transcription, like chromatin remodeling and termination, and importantly also mRNA processing and mRNP formation: Prp40, a subunit of the spliceosome, was shown to be recruited to hyper-phosphorylated repeats (Morris 2000). Npl3, as mentioned above, involved in splicing, 3' end processing and mRNA export, binds to the Ser2 phosphorylated CTD and is subsequently transferred to the nascent mRNA (Dermody et al. 2008).

Finally, increasing phosphorylation levels of Ser2, depletion of Ser5 and *app.* at cleavage and polyadenylation site declining Tyr1 phosphorylation levels mark the next transition: the recruitment of 3' end processing factors: Pcf11, a subunit of the cleavage factor IA (CF IA) binds to Ser2 phosphorylated CTD (Noble et al. 2005, Lunde et al. 2010) but not to Tyr1 and Ser2 di-phosphorylated repeats (Mayer et al. 2012a). Additionally, other 3' end processing factors bind to the Ser2 phospho-CTD: Rna14 (CFIA), Rna15 (CFIA), Ydh1 (CPF) and Yhh1 (CPF) (Barilla et al. 2001, Kyburz et al. 2003, Dichtl et al. 2002). Finally, upon termination, the CTD Ser2 phosphorylations are removed by the phosphatases Fcp1 and Ssu72 (Cho et al. 2001).

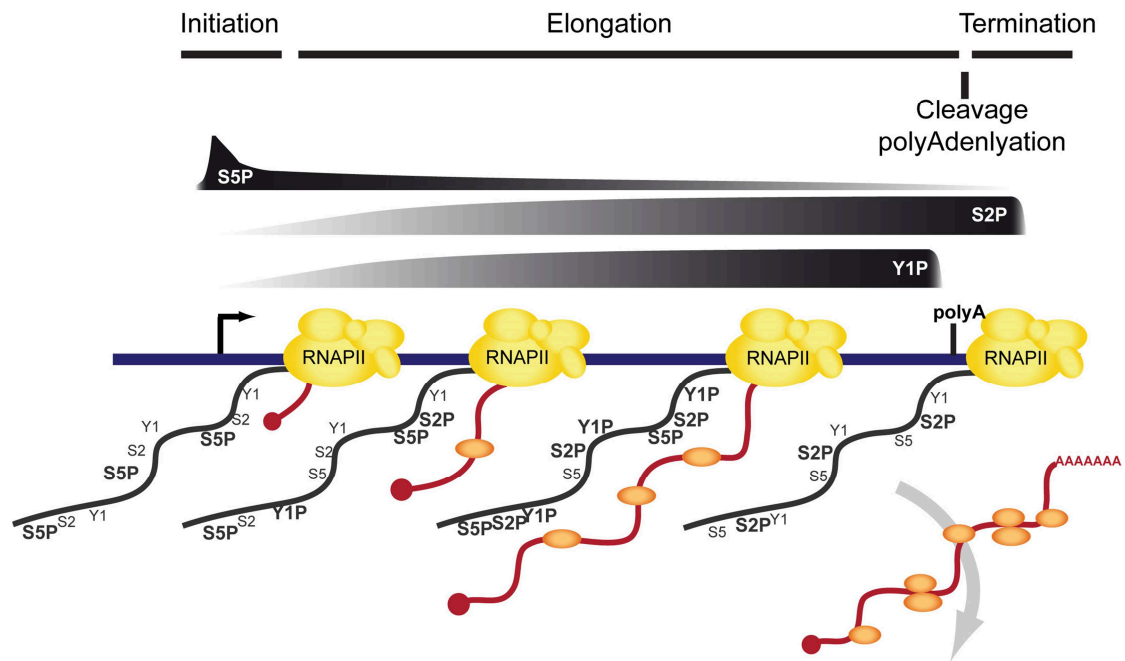


Figure 4: Phosphorylation of the Rpb1-CTD during the transcription cycle

The upper panel indicates the three main phases of the transcription cycle. Below the differential phosphorylation pattern of the CTD is depicted, as it can be found by ChIP-chip assays. In the lowest panel, an exemplary gene is shown with four RNAPII traversing through it. The CTD is depicted in black with the according phosphorylations, and the nascent mRNA is shown in red.

Additionally, also other CTD residues can be phosphorylated, *e.g.* Thr4, which is phosphorylated by CDK9, is needed in higher eukaryotes for correct 3' end processing of histone genes (Hsin et al. 2011). Thr4 phosphorylation was shown to be necessary for efficient transcription elongation in mammalian cells and to occur later than the Ser2 phosphorylation to the 3' end of the gene, however, yeast mutants lacking the Thr4 in their CTD repeats are viable (Hintermair et al. 2012).

8.2 The RNA as recruitment platform

The RNA itself also offers the cell the possibility to recruit proteins to the site of transcription and to the mRNP. Several proteins bind directly to the RNA via a recognize sequence motif. Examples for this are the many proteins involved in the 3' end processing, which *e.g.* recognize the positioning element or the efficiency element (for review see Mandel et al. 2008). Also subcomplexes of the spliceosome recognize their targets by RNA motifs, although here not proteins, but the RNA components of the snRNPs recognize the sequence by base-pairing (Will and Luhrmann 2011).

8.3 Spt5 - an upcoming recruitment platform

The general elongation factor Spt5, which binds to the body of RNAPII (Klein et al. 2011, Martinez-Rucobo et al. 2011) seems also to have a function as a recruitment platform during transcription. Spt5 is highly conserved and can be found in all three kingdoms of life (Werner and Grohmann 2011). The human homolog of Spt5 was shown to affect transcription elongation (for review see Hartzog and Fu 2013). A recent mass spectrometry study revealed that many proteins involved in mRNP biogenesis copurify with Spt5 in yeast, among those were capping enzyme proteins, 3' end processing proteins and also the hnRNP Tho1 (Lindstrom et al. 2003). Similar to Rpb1, Spt5 also has a repetitive C-terminal region (Spt5 CTR), which in yeast consists of 16 hexa-repeats, which were shown to be phosphorylated on their Ser1 residue during transcription elongation by Bur1 kinase in yeast or P-TEFb in humans (Liu et al. 2009, Yamada et al. 2006, Zhou et al. 2009). The CTR of Spt5 is needed for efficient recruitment of the Cleavage Factor I (CFI) to the transcription site (Mayer et al. 2012b). Additionally, the Paf1 complex, which is implicated in several processes like chromatin remodeling and 3' end processing (for review on Paf1 see Jaehning 2010), was shown to need correct Spt5-CTR phosphorylation for efficient recruitment (Qiu et al. 2012b). Moreover, the capping enzyme was also shown to need the Spt5-CTR for stable recruitment (Lidschreiber et al. 2013). Although Spt5 was recently established as a new recruitment platform in transcription, which is needed for proper transcription and 3' end formation, it is not known yet, if the Spt5-CTR is also important for mRNA export.

8.4 Additional ways to couple steps in mRNP biogenesis

Furthermore, additional principles to couple successive steps to each other are used: Once a protein is recruited to the mRNP, it can also serve to recruit other proteins, which are needed for the next step of gene expression: *E.g.* once Yra1 is recruited during transcription, it recruits later on Mex67 and then leaves the mRNP just before mRNA export.

Another principle is that two or more recruitment platforms are combined: The CFI complex can interact with the Rpb1-CTD as well as with the Spt5-CTR (Lunde et al. 2010, Mayer et al. 2012b). Also the Paf1 complex can interact with Rpb1-CTD and Spt5-CTR (Qiu et al. 2012b) and Npl3 was shown to bind to the Rpb1-CTD and to be transferred to the RNA (Lei et al. 2001, Dermody et al. 2008). Interestingly, She2, a protein involved in transport of the mRNA in yeast to the bud-tip was recently found to bind co-transcriptionally to Spt5 and then – in case the RNA is targeted for the bud-tip – She2 binds to the zip-code element, a RNA motif, to mediate transport (Shen et al. 2010). In this case recruitment by Spt5 ensures that She2 is available, while the RNA binding motif ensures specificity.

9. mRNA export

Before the mature mRNA can be translated into protein in eukaryotes, the export competent mRNP has to be exported from the nucleus to the cytoplasm (Figure 5). The mRNP has to cross the nuclear envelope, which consist of a lipid double layer, by traversing through the nuclear pore complex (NPC). The NPC is a very large multi-protein complex (more than 60 MDa in yeast and 125 MDa in higher eukaryotes), which consists of several subcomplexes. The NPC is a cylindrical structure, comprised of eight spokes, which form a central channel. The NPC can be divided into three segments: the nuclear face with the basket, the central channel and the cytoplasmic face with the fibrils. *App.* 30 different nuclear pore proteins (Nups) constitute one NPC. Three Nups tunnel as transmembran proteins the nuclear envelope and form the outer ring of the NPC, which connects the NPC to the nuclear envelope. Additional scaffold Nups form the outer and inner ring of the NPC. This NPC scaffold is bound by several Nups containing phenylalanine and glycine repeats (FG repeats). These repeats appear to be largely unfolded and protrude into the inner channel of the NPC. The FG repeats have a pivotal role in all transport processes through the NPC. In particular, they serve as binding sites for mRNA export and other transport proteins (for review of the NPC see Oeffinger and Zenklusen 2012).

mRNA export itself is mediated by the essential export receptor Mex67-Mtr2. This heterodimer (Santos-Rosa et al. 1998) is highly conserved from yeast to human, where Mex67 is named NXF1 or TAP and Mtr2 p15/NXT1 (Strässer et al. 2002, Katahira J et al. 1999). The conservation is so high, that the lethal yeast knock out of *MEX67/MTR2* can be complemented by expression of the human homologs (Katahira et al. 1999). Interestingly, Mex67-Mtr2 can directly bind to RNA, but was shown to have only a very low affinity for direct binding to RNA itself (Katahira et al. 1999, Strässer and Hurt 2000). It was shown, that Mex67-Mtr2 is recruited to mRNAs by direct interaction with mRNA binding proteins, which serve as export receptor adaptors and increase the efficiency of mRNA export. One of those is Yra1 (in higher eukaryotes Aly) (Strässer and Hurt 2000, Katahira J et al. 1999), which is a subunit of the conserved TREX complex. Moreover, Hpr1, Nab2 and Npl3 are also implicated in recruiting the export receptor to the mRNP (as described above). Mex67-Mtr2/TAP-p15 was shown to shuttle together with the mRNP from the nucleus to the cytoplasm and seems to be mainly localized at the inner nuclear pore in yeast and higher eukaryotes (Katahira et al. 1999, Rout et al. 2000). Mex67 interacts with the NPC's FG repeats via its ubiquitin associated domain (Katahira et al. 1999, Strasser et al. 2000, Hobeika et al. 2009). On the cytoplasmic site a remodeling step of the mRNP takes place in which several proteins are released from the mRNP, to promote directionality of mRNA export. This remodeling step is brought about by the ATP-dependent DEAD box helicase Dbp5, which binds to the mRNP in the nucleus and shuttles together with it through the

NPC (Linder and Jankowsky 2011). At the cytoplasmic face of the NPC the nuclear pore protein Gle1 and Inositol hexakisphosphate (IP_6) activate Dbp5, which then remodels the mRNP (Tran et al. 2007, Noble et al. 2011, Alcázar-Román et al. 2006). This leads to the release of several proteins involved in mRNA export, *e.g.* Mex67-Mtr2 and Nab2 (Lund and Guthrie 2005 and Tran et al. 2007), and the mRNA is released into the cytoplasm for translation.

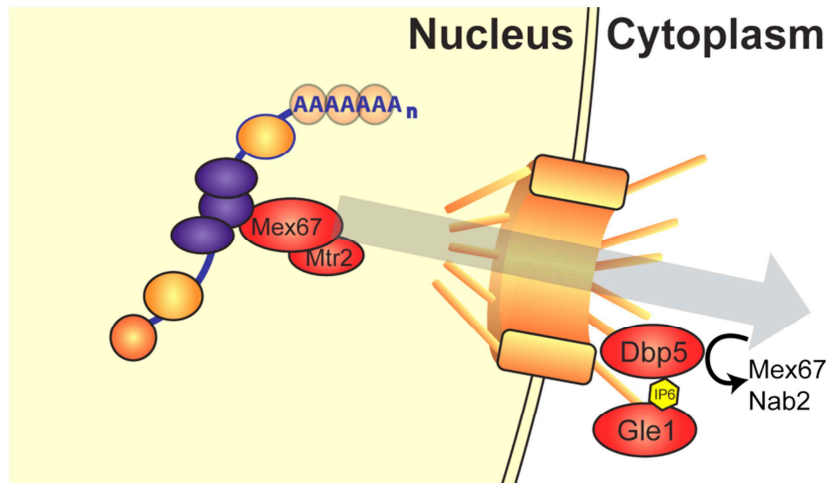


Figure 5: mRNA export through the Nuclear Pore complex.

The export competent mRNP is exported by Mex67-Mtr2 (red) through the NPC (orange) to the cytoplasm. The transport is driven by a Mex67/Mtr2 gradient: on the cytoplasmic face of the NPC, Mex67 is dissociated from the mRNP by the DEAD box helicase Dbp5, which is activated by IP_6 (yellow) and the NPC bound Gle1 (red).

10. Aims and scope of this study

The TREX complex is a key player in early mRNP formation and mRNA export. It has a variety of functions spanning from transcription to mRNA export, but until now most data is based on single gene data. Whether TREX, Nab2 and Npl3 are recruited to all genes remains as an important open question. Furthermore, it is not known if there are several mRNA export pathways or only one canonical.

Although for the last decade several studies investigated THO/TREX function and recruitment, it is still an important open question how THO/TREX is recruited on a molecular basis. Recently Prp19 was shown to stabilize THO/TREX occupancy at the 3' end of the genes by the interaction of the Prp19 complex with THO/TREX and RNAPII (Chanarat et al. 2011), but the recruitment mechanism of TREX at the 5' end, remained unknown. Furthermore, Yra1, an essential TREX subunit, which directly binds to the RNA and the export receptor Mex67-Mtr2, was shown to be not only recruited by TREX, but also by Pcf11, independently of other TREX subunits (Johnson et al. 2009, Johnson et al. 2011). Moreover, Yra1 was recently also shown to bind directly to the phosphorylated CTD of the RNAPII

(MacKellar and Greenleaf 2011). Interestingly, all these interactions were mapped to the same region, were Yra1 also interacts with Mex67 and Sub2 and thereby TREX, indicating that these recruitment mechanisms might be mutually exclusive or alternative to each other. However, the fact that deletion of this interaction region is not lethal for the cell and Yra1 completely mislocalizes to the cytoplasm indicates that all these interactions are at least not the only mechanism to recruit the mRNA export machinery to the site of transcription. Briefly, how TREX is recruited on a molecular level is still unknown.

Furthermore, it is a fundamental question how the two major recruitment platforms of transcription, the nascent RNA and the Rpb1-CTD, are involved in the mRNA export and if the newly found Spt5-CTR also has a function in mRNP formation and export by regulating the recruitment of TREX.

Additionally, given the diverse functions and the highly dynamic character of the TREX complex, it is of special interest, whether the whole TREX complex is always recruited as one complex or whether there are different compositions, *e.g.* a complex which is needed to preserve genes with a special GC content from elevated transcription dependent hyper-recombination frequency.

To answer these questions on a genome-wide basis, ChIP-chip data of the THO/TREX complex, Tho1, Nab2 and Npl3 were generated. Bioinformatics analysis should answer the question, whether the mRNA export factors are recruited to all protein coding genes, if factors are recruited to non-protein coding genes and if there are different mRNA export pathways present in the cell, which are specific for the different mRNA export receptor adaptors. In a next step, the data should be analyzed to learn more about the recruitment mechanism of the factors. By combining both, genome-wide recruitment data and classical biochemical analysis, the molecular recruitment mechanism of THO/TREX and the role of the Rpb1-CTD and Spt5-CTR in mRNA export should be elucidated.

II. Results

1. TREX recruitment

It is known for several years that the TREX complex is one of the key players in nuclear mRNA export. But by now the genome-wide function of TREX has not been investigated in depth. To analyze the genome-wide function of TREX we determined its recruitment to genes by ChIP-chip. Before ChIP-chip experiments were carried out, all strains were tested for correct expression of the TAP-tagged proteins by Western blotting and TAP purification. To reach a high resolution in the ChIP-chip experiments, the DNA fragments were sheared to an average fragment size of *app.* 200-250 bp and hybridized to tiling arrays with a four nucleotide resolution throughout the whole yeast genome (David et al. 2006). Each protein was assayed in two to four independent biological replicates, dependent on the Pearson correlation coefficient of the raw data of the replicates. For data analysis we used a normalization procedure that corrects for unspecific binding and GC content of the array probes by using mock IP and input measurements, based on (Mayer et al. 2010), but with small modifications: single probes, which did not correlate well between the replicates, were omitted and interpolated by the neighboring and overlapping probes. We transformed the data into linear occupancies, as in (Mayer et al. 2010), setting the 10% quantile to 0 and the 99.8% quantile to 1. Some of the ChIP-chip experiments have been performed by Cornelia Burkert-Kautzsch.

First, we determined the occupancies of Rpb3-TAP, a TAP tagged core subunit of RNAPII. As expected and published earlier by other labs (Mayer et al. 2010), Rpb3 was recruited to protein coding genes and sn/snoRNAs as well as to loci producing cryptic unstable transcripts (CUTs) and small unannotated transcripts (SUTs) as defined by (Xu et al. 2009). Consistently, we did not observe any binding to RNAPI and RNAPIII loci. We next calculated meta profiles for Rpb3. Therefore all genes of a certain length class were selected. The gene classes used here were previously described in (Mayer et al. 2010). Only 'verified' or 'uncharacterized' genes annotated in the *Saccharomyces* Genome Database) with annotated Transcription Start Site (TSS) and polyA sites (pAs) were considered. We considered the top 50% expressed genes, according to the measurements by (Dengl et al. 2009) for the meta profile calculation. The Meta profiles in each group were scaled to the median length and the meta profiles were calculated by taking the mean over the scaled profiles. Rpb3 profiles showed the typical polymerase profiles, with a big peak at *app.* initiation site of transcription and a second, smaller peak at *app.* the transcription termination site (Figure 6).

In the next step we determined the meta profiles of the THO complex subunits Hpr1, Tho2 and Mft1 with TAP-tagged yeast strains. All profiles of the different subunits of the stable THO complex had a very high correlation among each other, as expected for the formation of one stable complex (Figure

15). We did not observe substantial amounts of THO at RNAPII or RNAPIII genes, but recruitment to all RNAPII transcribed genes: protein coding as well as sn/snoRNA, CUT and SUT genes. The calculated meta occupancy profiles for the THO subunits to protein coding genes clearly showed, that THO is recruited during elongation phase of transcription, consistent with earlier single gene analysis (Abruzzi et al. 2004) and genome wide analysis of Hpr1 (Gómez-González et al. 2011). As indicated by very high correlation coefficients between the THO complex subunits and the almost identical profiles, THO seems to be recruited as one complex during elongation phase. The occupancy of THO was increasing from 5' to 3' of the gene during transcription elongation. As expected for an mRNA export factor, the ChIP signals dropped at *app.* the polyadenylation and cleavage site (Figure 6).

Yra1 and Sub2, which bind to the THO subcomplex to form TREX, showed similar meta profiles as the THO subunits. Yra1 and Sub2 were also recruited during elongation phase, their meta profiles increased similar to THO from 5' to 3' and the proteins left the chromatin at *app.* the polyadenylation site, slightly upstream of the THO complex (Figure 6). Furthermore, Yra1 and Sub2 Pearson correlation coefficients to the THO subunits were very high (Figure 15), indicating that Yra1 and Sub2 are most likely recruited to all genes, where THO is also recruited.

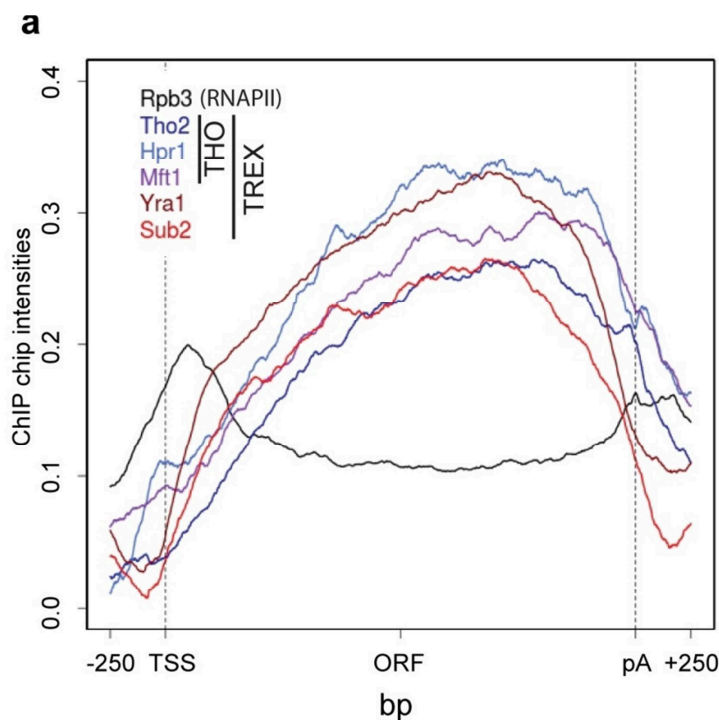


Figure 6: TREX recruitment increases from the 5' to the 3' end of the gene

Meta gene occupancy profiles of the TREX components Tho2, Hpr1, Mft1, Sub2 and Yra1. For comparison the meta gene occupancy profile of RNAPII (Rpb3) is shown. The 50% highest genes (299 genes) from 1538 to 2895 bp were used for calculation.

Interestingly, the two TREX subunits Hrb1 and Gbp2, showed different profiles: Both SR proteins were recruited during the elongation phase, but did not increase like THO, Sub2 and Yra1 from 5' to 3', but they seem to reach a plateau throughout the body of the gene after 400 bp (Figure 7), which slightly declined towards the 3' end. The different shape of the profiles of Gbp2 and Hrb1 resulted in a lower correlation coefficient of the two SR proteins to the other TREX subunits (Figure 15). However, we found Gbp2 and Hrb1 at all genes, at which THO was present. This is unexpected, since Gbp2 and Hrb1 are TREX subunits and were expected to show the same occupancy profiles as THO/TREX. The difference could hypothetically be explained by a mechanism in which all TREX subunits are recruited together and then Gbp2 and Hrb1 might dissociate from the other TREX subunits.

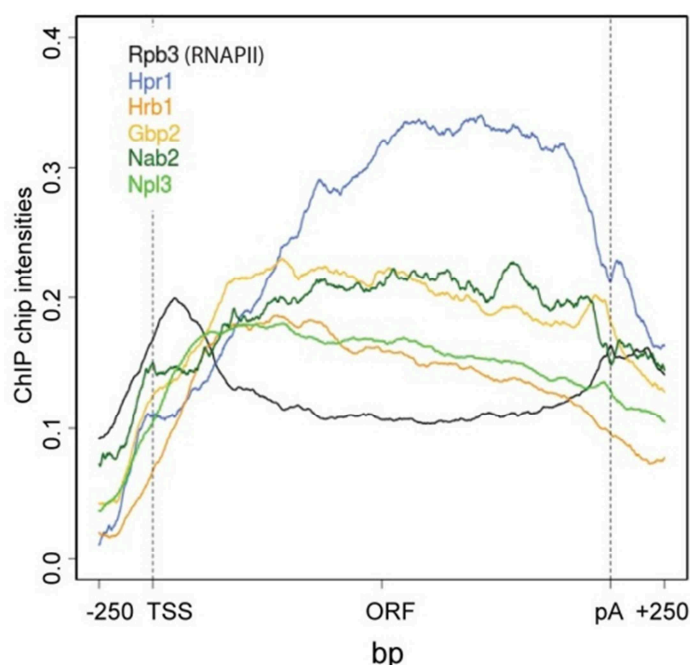


Figure 7: Meta profiles of Hrb1, Gbp2, Nab2 and Npl3 are distinct different from THO, Sub2 and Yra1.

Meta gene occupancy profiles of the TREX components Gbp2 and Hrb1 and the mRNA-binding proteins Nab2 and Npl3. The occupancy of these mRNA-binding proteins reaches a plateau in the open reading frame, which slightly declines from 5' to 3' of the gene. For comparison the meta gene occupancy profiles of RNAPII (Rpb3) and the TREX component Hpr1 are shown.

We next excluded possible effects due to the C-terminal TAP-tag on Gbp2 which could impair recruitment. Therefore we compared the recruitment of N- and C-terminal TAP-tagged Gbp2 strains on single genes and did not detect any significant differences in the recruitment profiles between C-

and N-terminal TAP-tagged strains (Figure 8), ascertaining that the profile of Gbp2 is not influenced by the C-terminal TAP fusion. This experiment was performed by Cornelia Burkert-Kautzsch.

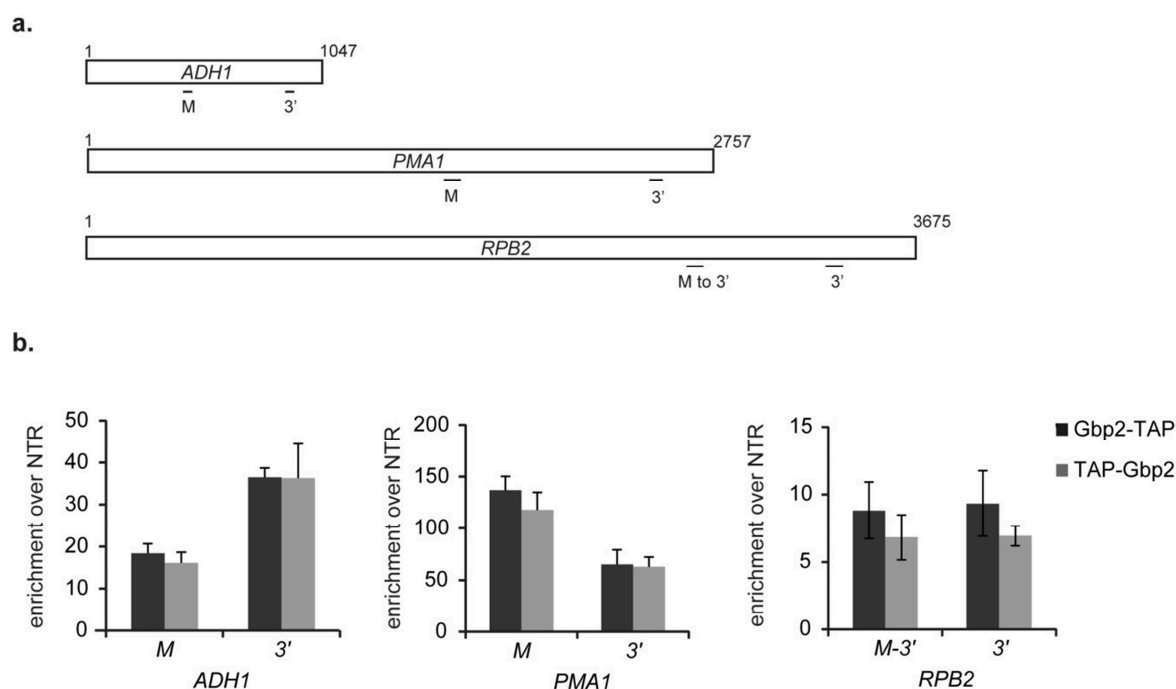


Figure 8: Gbp2-TAP and TAP-Gbp2 occupancy is indistinguishable from each other

(a) Primer positions of *ADH1*, *PMA1* and *RPB2* are depicted and drawn to scale. For *ADH1* primer pair M amplifies nts 408-476 and primer pair 3' nts 916-966. For *PMA1* the primer pair M amplifies nucleotides nts 1574-1651 and the primer pair 3' amplifies nts 2484-2543. For *RPB2* the primer pair M to 3' amplifies nucleotides nts 2624-2704 and the primer pair 3' amplifies nts 3288-3361. **(b)** ChIP results for Gbp2-TAP and TAP-Gbp2. Independent of the N- or C-terminal TAP-tag fusion to Gbp2, ChIP enrichments on *ADH1*, *PMA1* and *RPB2* showed the same results and are indistinguishable from each other, indicating that the TAP-tag did not interfere with Gbp2 recruitment. These experiments have been performed and analyzed by Cornelia Burkert-Kautzsch.

Additionally, the recruitment profiles of the mRNA export proteins Nab2 and Npl3 were determined. Nab2 and Npl3 are, as described above, also export receptor adapters and have various functions in gene expression. ChIP-chip showed that both proteins are recruited during elongation phase of transcription and leave the site of transcription at the polyadenylation and cleavage site, similar as THO/TREX (Figure 7). This is consistent with a function in mRNA export and also reflects the role of Nab2 and Npl3 in 3' end processing (Hector et al. 2002 and Bucheli and Buratowski 2005). Recruitment of both proteins reached a plateau in the gene body, which was slightly decreasing towards the 3' end. Interestingly, these profiles resembled very much the Hrb1 and Gbp2 profiles and also the Pearson correlation coefficients of Hrb1 and Gbp2 with Nab2 and Npl3 were higher than with THO/TREX (Figure 15).

2. TREX, Nab2 and Npl3 recruitment to noncoding RNAs

It was not known, whether TREX, Nab2 and Npl3 are recruited to non-protein coding genes in addition to protein coding genes. Thus, we inspected the genome wide data for recruitment of TREX, Nab2 and Npl3 to non RNAPII genes. We found all mRNA export proteins recruited to sn/snoRNA genes. However, TREX occupancy at those genes was considerable smaller as expected by the high transcriptional frequency (Rpb3 occupancy) of sn/snoRNA genes. Furthermore, Sub2 and also Yra1 were recruited in minor amounts to RNAPIII genes like tRNAs, but we did not find THO subunits at tRNA genes (Figure 9). However, Yra1 and Sub2 levels were much lower at RNAPIII genes compared to RNAPII genes, although RNAPIII genes are transcribed with a much higher frequency, indicating only a minor function in RNAPIII transcription if any. These results are also in line with the recently published data by (Johnson et al. 2011), which also found modest levels of Yra1 and Sub2 at tRNA genes.

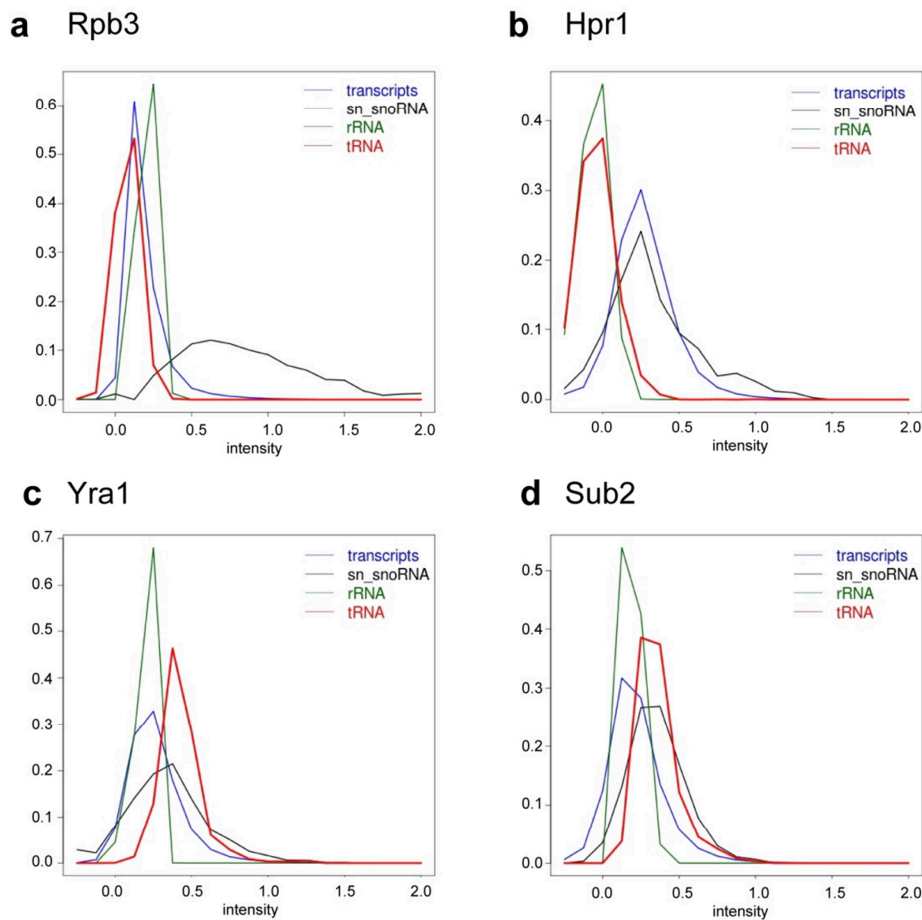


Figure 9: Rpb3, Hpr1, Yra1 and Sub2 are recruited to protein coding and only in minor amounts to sn/snoRNA genes

The occupancy for the probes annotated to each feature (transcripts (protein coding genes), sn/snoRNA genes, rRNA (ribosomal RNA genes) and tRNA genes) was plotted in a density plot for (a) RNAPIII (Rpb3), (b) Hpr1 (THO), (c) Yra1 (TREX) and (d) Sub2 (TREX). Hpr1, Yra1 and Sub2 showed lower occupancy at sn/snoRNA genes as expected by the high occupancy on those genes by RNAPII (Rpb3). Additionally, Yra1 and Sub2 show occupancies at tRNA genes, indicating that Yra1 and Sub2 are recruited to tRNA genes.

Furthermore, Nab2 was found to be heavily recruited to RNAPIII transcribed loci, but not to RNAPI loci (Figure 10, a). In the meta profiles of RNAPIII (Rpc160) and Nab2 for tRNA genes, strong recruitment could be seen for both factors (Figure 10 b-d), both RNAPIII (Rpc160) and Nab2 have similar profiles for tRNA genes, while RNAPII (Rpb3) was not detectable. Furthermore, also other RNAPIII genes, *e.g.* *SCR1*, showed high Nab2 occupancy (Figure 10 c). Another indication for Nab2 carrying out a function in RNAPIII RNA metabolism is, that Nab2 occupancy not only at RNAPII genes had very high Pearson correlation coefficients with RNAPII (Rpb3) recruitment, but also Nab2 and RNAPIII (Rpc160) occupancies were highly correlated at tRNA genes, while RNAPIII (Rpc160) and Nab2 occupancies on RNAPII genes as well as RNAPII (Rpb3) and Nab2 occupancies at RNAPIII genes did not correlate well (Figure 10e, f).

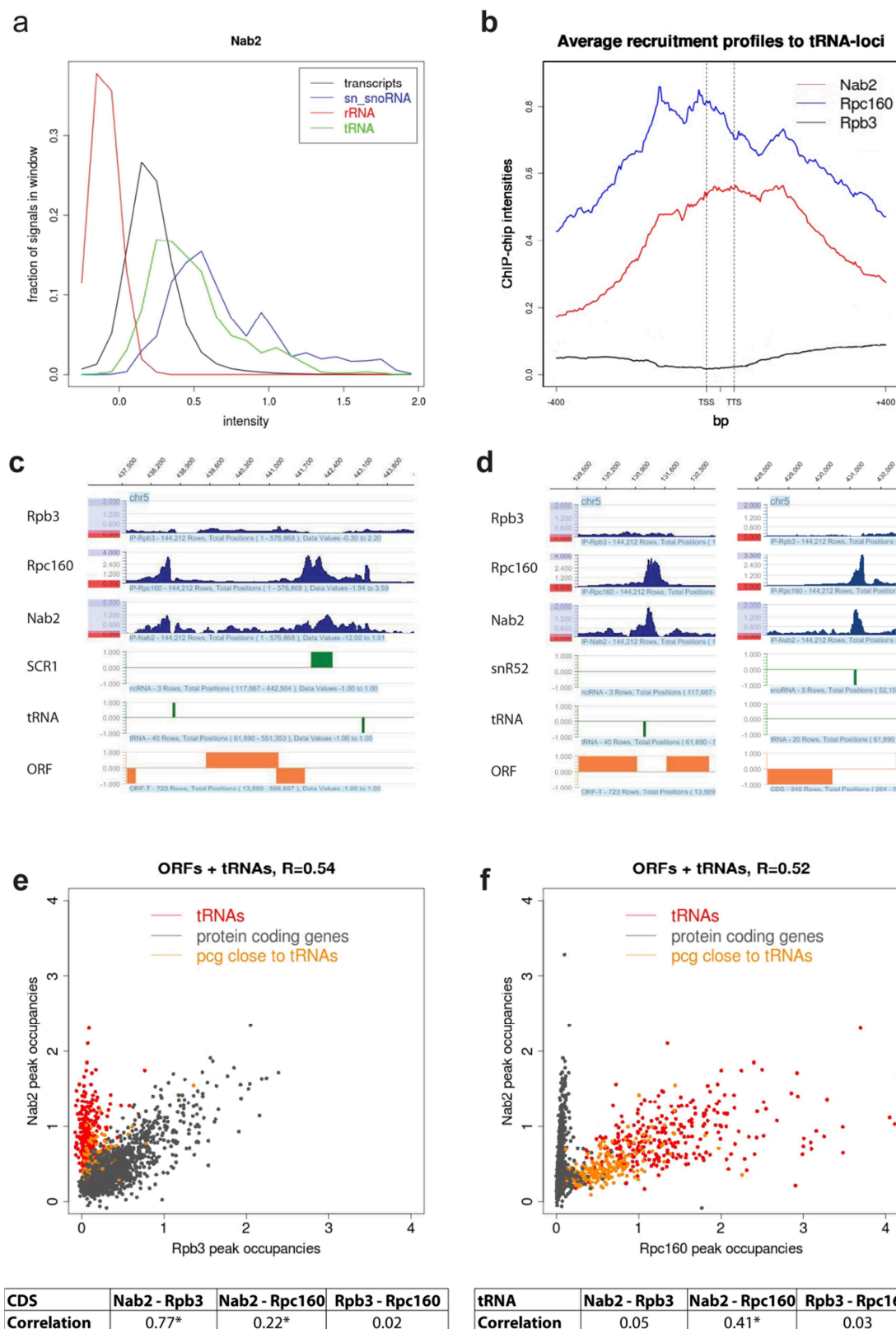


Figure 10: Nab2 is recruited to RNAPIII transcribed loci and correlates with RNAPIII at tRNA loci

(a) Density plots of the signals for the four classes: transcripts (protein coding genes), sn/snoRNA, rRNA (ribosomal RNA) and tRNA for Nab2. Nab2 occupancy on tRNA genes is even higher than on protein coding genes. (b) Meta gene occupancy profiles for tRNA genes for RNAPII (Rpb3), RNAPIII (Rpc160) and Nab2. (c) Recruitment for RNAPII (Rpb3), RNAPIII (Rpc160) and Nab2 to the RNAPIII gene *SCR1* and to two tRNA genes on Chromosome V. (d) As c but for the RNAPIII gene *snR52* and a tRNA. (e) Plotted are the peak occupancies for Nab2 and Rpb3. The grey dots represent protein coding genes which are at

least 250 bp separated from tRNA genes. tRNA genes are plotted in red. RNAPII genes which neighbor within 250bp a tRNA gene are depicted in orange. Due to the high signals caused by the tRNA genes, which spill over to the RNAPII genes, is the correlation coefficient of those genes to RNAPII worse than the correlation coefficient to RNAPIII. Pearson correlation coefficients for protein coding genes are given in the panel below. (f) As e, but for Nab2 and RNAPIII (Rpc160). The panel below gives the Pearson correlation coefficients for tRNA genes.

Recently, Tuck and Tollervey (2013) published a genome-wide CRAC dataset for Nab2 binding to RNA. Our examination of this dataset showed, that Nab2 is associated *in vivo* with tRNAs and other RNAPIII transcripts like *e.g.* *SCR1* and *RPR1*. Interestingly, the mapping of the bound RNA to the yeast genome revealed that Nab2 was not only bound to the mature forms, but also up- and downstream of the ncRNAs (*e.g.* for *RPR1* and with lower signals downstream of many tRNAs and *SCR1*), showing very early binding to the immature forms of RNAPIII transcripts (Figure 11). This result supports the observation from our ChIP-chip datasets and indicates that Nab2 might be involved in RNAPIII transcription or early transcript biogenesis.

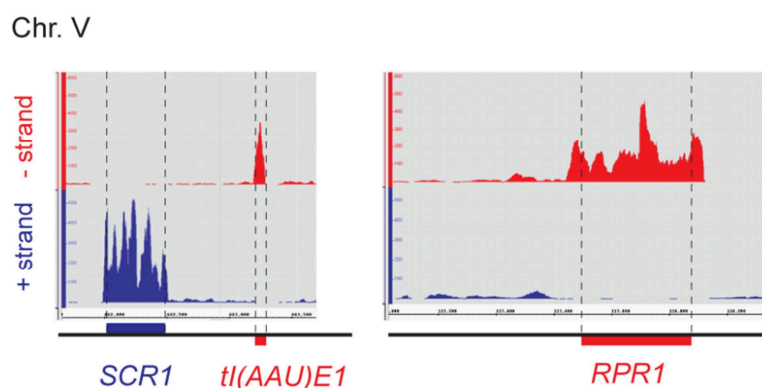


Figure 11: Nab2 binds *in vivo* to immature RNAPIII transcripts

CRAC data for Nab2 from (Tuck and Tollervey 2013) was analyzed for binding to RNAPIII transcripts. The CRAC profiles recapitulate the binding profiles found by ChIP-chip and Nab2 can clearly be found in the early, immature transcript of *RPR1* (right panel) indicating a co-transcriptional binding event of Nab2 to the RNAPIII transcripts. For *SCR1* and *tI(AAU)E1* almost all signal maps to the mature transcript, however, a low signal can be observed in the downstream region. The plus strand is depicted in blue, the minus strand in red. The lowest panel depicts the according genes as annotated by SGD in the color of the according strand.

3. THO/TREX is recruited length dependent

To answer the long standing question, whether THO/TREX has specificity for special classes of genes, THO/TREX recruitment was analyzed for any specificity towards special classes of genes. We found that THO/TREX, Nab2 and Npl3 were recruited to all transcribed genes, though to changing amounts relative to RNAPII. Firstly, we observed, that Yra1 is recruited with a slight preference to non-intron containing genes. When the recruitment of TREX (Hpr1 and Yra1, Figure 12 d and e) to intron containing genes was compared to other general recruited factors like the RNAPII (Rpb3, Figure 12 a)

or the elongation factor Spt6 (Figure 12 b) or Nab2 (Figure 12 c) it is striking, that TREX occupancy at intron containing genes was lower as expected, but TREX was still present at those genes (Figure 9 d-e). Furthermore, Npl3 was recruited to higher levels to intron-containing genes (Figure 12 f). This is in line with a possible function of Npl3 in promoting the recruitment of the spliceosome to introns (Kress et al. 2008).

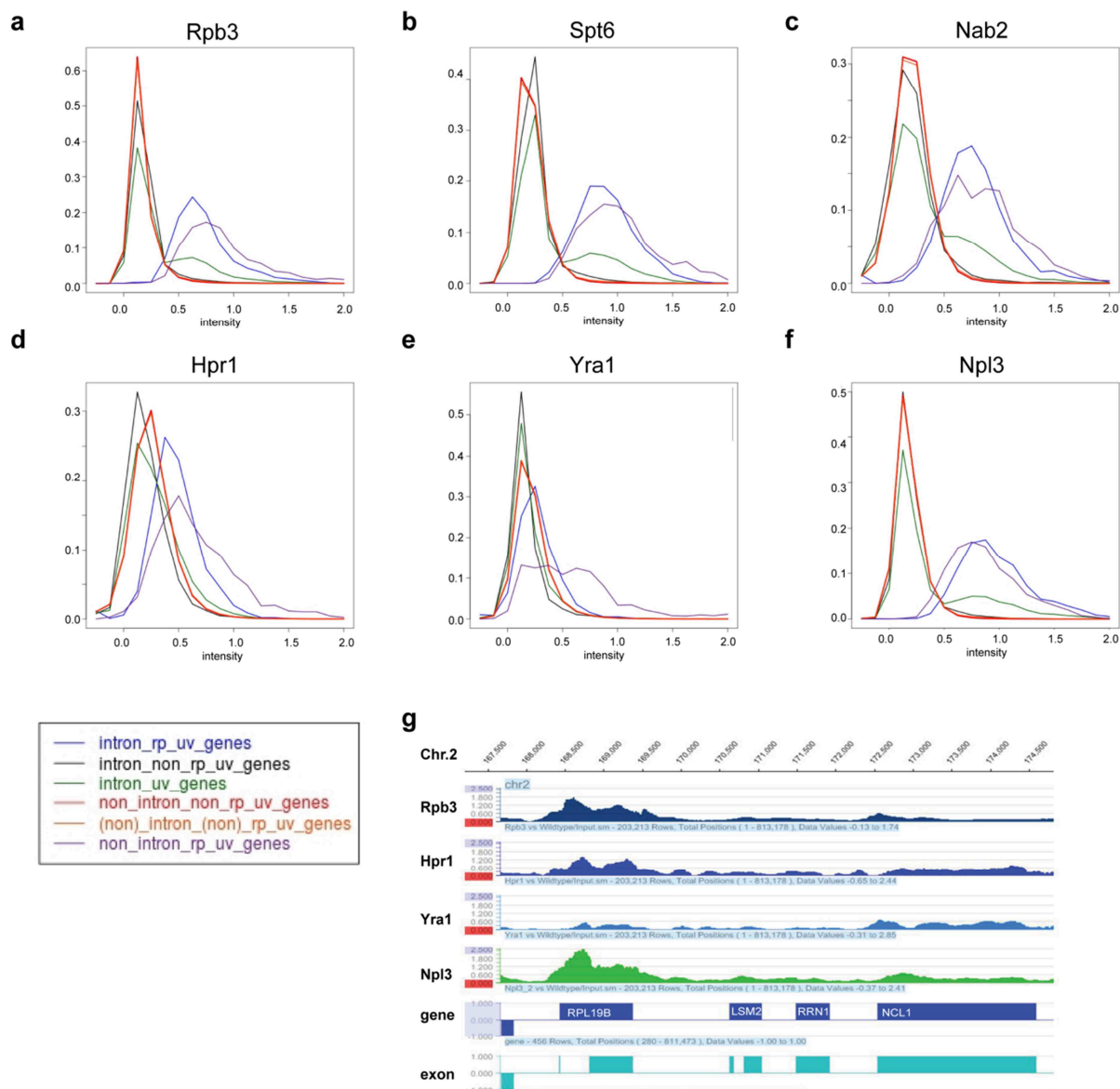


Figure 12: Yra1 binds preferentially to non-intron containing genes, Npl3 is recruited to higher levels to intron containing genes.

(a)-(f): Density plots of RNAPII (Rpb3), the general elongation factor Spt6, Nab2, TREX (Hpr1 and Yra1) and Npl3. The occupancy for the probes annotated to each feature as plotted in a density plot. Ribosomal protein coding genes containing introns are depicted in blue, non-ribosomal protein coding genes with introns are black; all intron containing genes are green. Non-intron containing non-ribosomal protein coding genes are shown in red and orange, non-intron containing ribosomal protein coding genes are depicted in magenta. **(g)** Occupancies of RNAPII (Rpb3), Hpr1 (TREX), Yra1 (TREX) and Npl3 on a region of Chromosome 2. Importantly, the occupancy at *NCL1* (rightmost gene) of Hpr1, Yra1 and Npl3 is equivalent high, but Npl3 occupancy is much higher at *Rpl19B*.

To analyze whether the occupancy constantly increases with growing gene length or whether the occupancy at some point reaches a maximum plateau, we set up 8 length groups for genes between 500 and 5000 bp. For each gene we calculated the peak occupancy of each factor, which represents the 90% quantile of the occupancy and gives a robust measure for the maximum recruitment to each gene. To correct for the different transcriptional intensities of the genes, we divided each peak occupancy by the according peak occupancy of RNAPII (Rpb3). The values for each gene class were averaged and plotted versus the class mean length. As expected by the meta profile plots for different length classes (Figure 14), the plots for the length dependency showed a striking increase of the peak occupancies of THO, Sub2 and Yra1 with increasing gene length. The increase was very steep up to around 2000 bp and then flattens slightly, but persists for genes up to 5000 bp length (Figure 13). Noteworthy, the THO, Sub2 and Yra1 meta profiles for the different length classes were perfectly superimposed, until they reach the polyA site of each length class, where the signal drops (Figure 14).

Contrary to the other TREX subunits, Hrb1 and Gbp2 did not show such a 5' to 3' increase over the whole gene length. Also Nab2 and Npl3 did only show a very slight increase over the gene length and were therefore considered as not length dependent. The difference of length dependent recruitment and a stable maximum recruitment could also be seen in the correlations of the peak occupancies for the proteins: THO, Sub2 and Yra1 correlated very well among each other, while Nab2, Npl3, Hrb1 and Gbp2 correlated less with THO, but very well among each other and with general elongation factors that were also not length dependent (Figure 16).

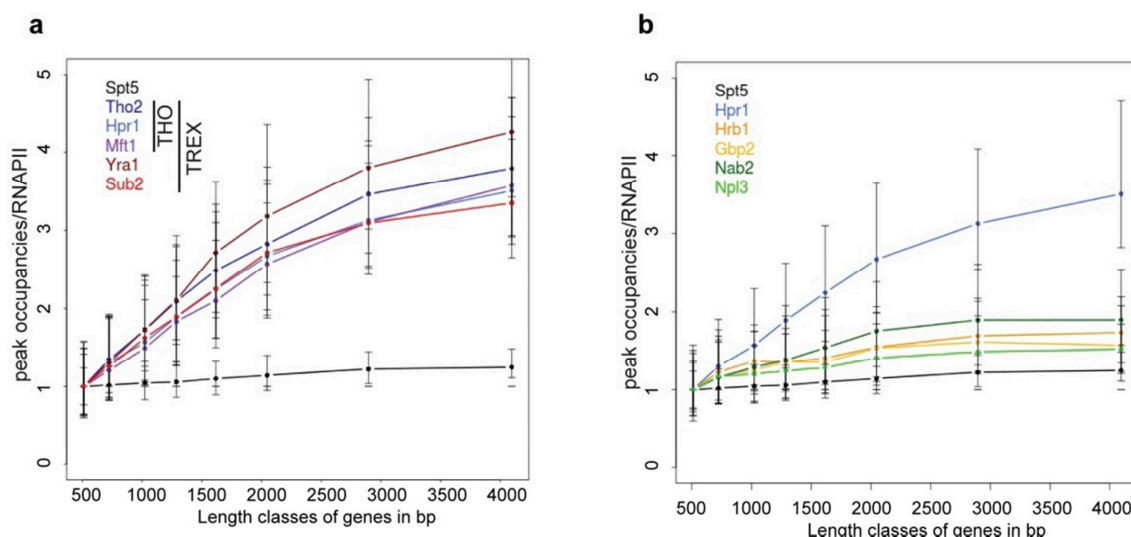


Figure 13: TREX is recruited length dependent to the genes

(a) THO, Sub2 and Yra1 peak occupancies increase with gene length. Genes were subdivided according to the indicated length classes and their occupancy was normalized. Length classes are: A (512-723bp), B (724-1023 bp), C (1024-1286 bp), D (1287-1617 bp), E (1618-2047 bp), F (2048-2895 bp), G (2896-4095 bp), H (4096-5793 bp). **(b)** Occupancy of mRNA-binding proteins Hrb1 (TREX), Gbp2 (TREX), Nab2 and Npl3 increase only slightly with gene length. Graph as in (a).

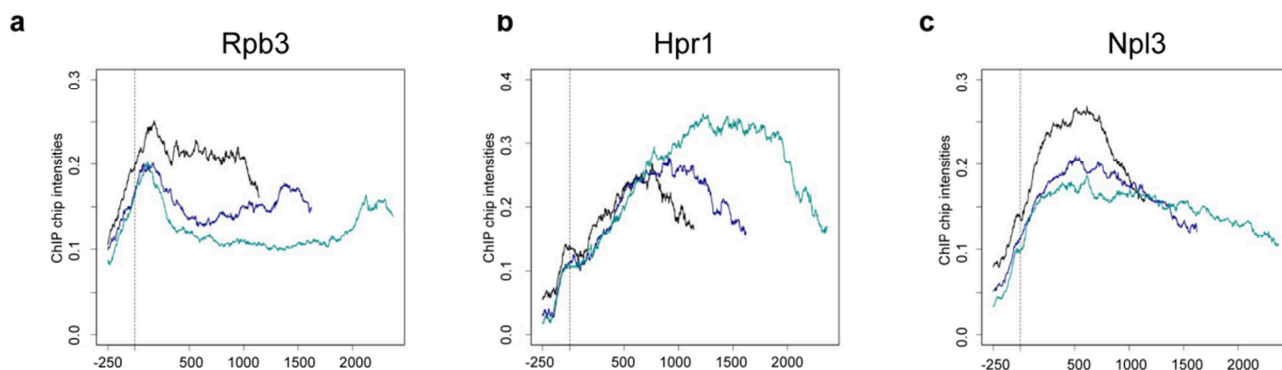


Figure 14: THO/TREX occupancy increases throughout the ORF

Meta profiles for short, middle and long genes aligned to the Transcription Start Site. **(a)** RNAPII (Rpb3) recruitment is not length dependent, **(b)** Hpr1-TAP (THO/TREX) increases from 5' to 3'. **(c)** Npl3 recruitment is also not length dependent.

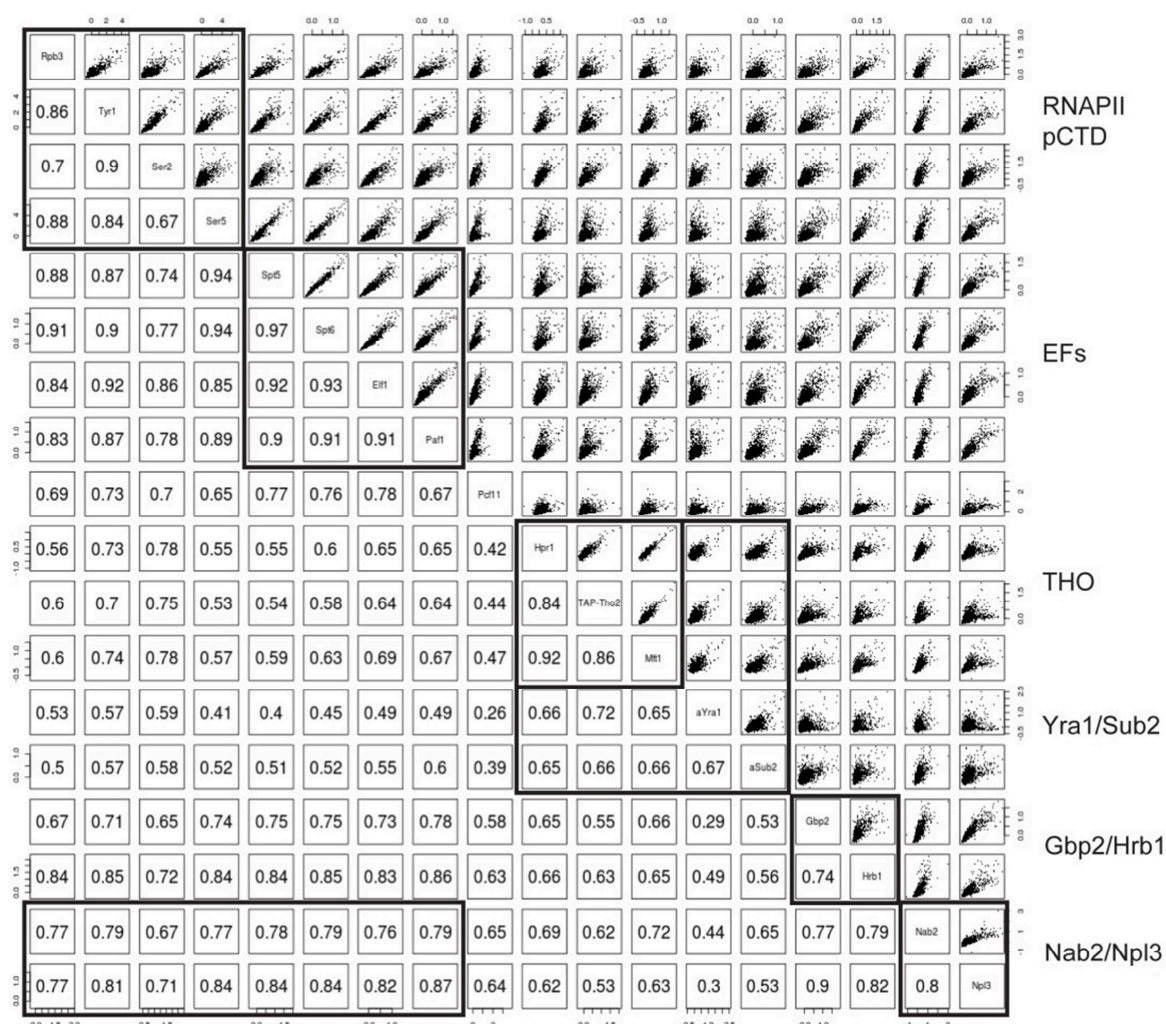


Figure 15: Pearson Correlation coefficients between TREX components, Nab2, Npl3 and general elongation factors.

For each protein the peak occupancies for each gene (90th percentile of each profile for each gene) were calculated. The numbers give the Pearson correlation coefficient of two proteins, in the graph the black points represent the peak occupancies. As expected, the general elongation factors Spt5, Spt6 and Elf1 correlate very highly with each other. Also RNAPII (Rpb3) and the phospho-CTD marks Y1P, S2P and S5P correlate very well with each other and with the general elongation factors. S2P correlates – due to the length dependency – less well with general elongation factors, but highly with THO subunits. As expected, the THO subunits, Sub2 and Yra1 correlate highly. Gbp2 and Hrb1 correlate less with THO subunits, Sub2 and Yra1, but highly with Nab2 and Npl3.

We also analyzed several available, comparable ChIP-chip data sets of *bona fide* general elongation factors and proteins involved in mRNP biogenesis for their length dependency (datasets from: Mayer et al. 2010 and Mayer et al. 2012b). Strikingly, we only found pronounced length dependency for THO, Sub2 and Yra1 but not for *bona fide* elongation factors like Spt5, Spt6, Spt16, Bur1 and Paf1 (Figure 16). This was also reflected by a good correlation coefficient between the non-length dependent factors.

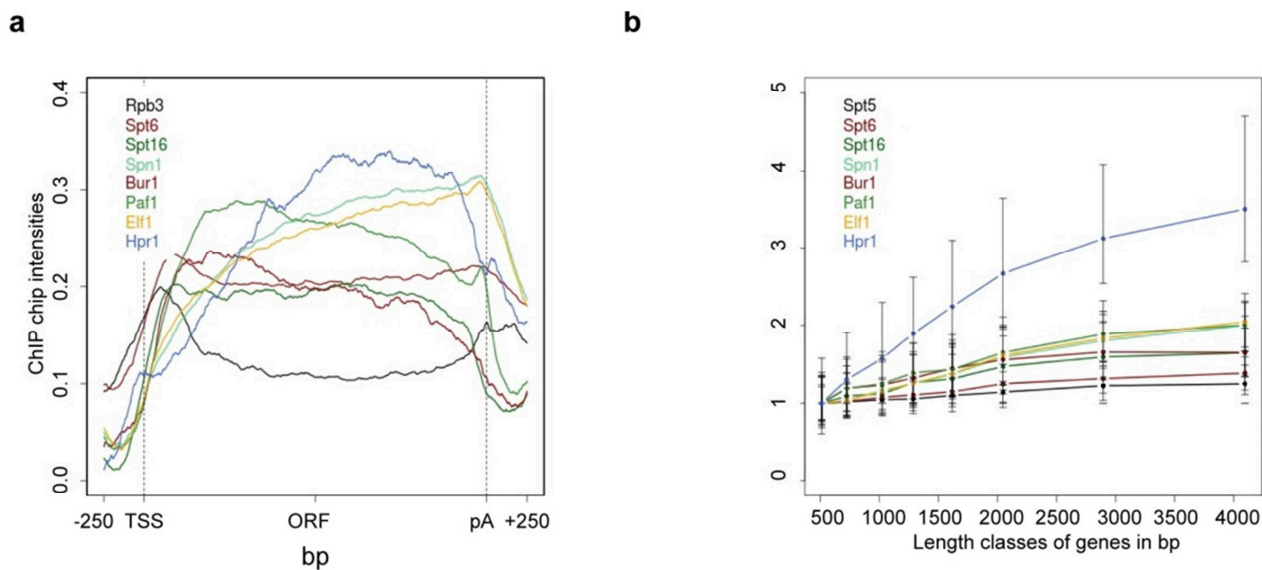


Figure 16: *bona fide* general elongation factors are not length dependent.

(a) Meta gene occupancy profiles and (b) peak occupancy according to length classes for RNAPII (Rpb3) (only a), the transcription elongation factors Spt5 (only b), Spt6, Spt16, Bur1 and Paf1 and the TREX component Hpr1.

In summary, the core TREX complex (THO, Sub2 and Yra1) is recruited to all protein coding genes during elongation phase of transcription. Most likely, all THO subunits, Yra1 and Sub2 are recruited as one core complex, as shown by very high correlation between the occupancies of these proteins. Remarkably, occupancies of THO, Sub2 and Yra1 increase almost linearly from 5' to 3' with growing gene length. This increase seems to be unique for THO, Yra1 and Sub2, since we did not observe any similar effect for general RNAPII elongation factors. Interestingly, Gbp2, Hrb1, Nab2 and Npl3 showed no length dependency and therefore pointed towards a different recruitment mechanism or dissociation from the transcription site after recruitment. Hrb1 and Gbp2 were recruited to all genes, which recruit TREX, but occupancy profiles were distinct different in their shape, possibly indicating dissociation of Gbp2 and Hrb1 after the recruitment of the whole TREX complex to the transcription site.

4. TREX recruitment is RNA dependent

The length dependent recruitment of THO, Sub2 and Yra1 to protein coding genes seemed to be unique, since we did not find a similar increasing occupancy for any other transcription elongation factor. We hypothesized, that such a 5' to 3' increase could be mediated by several similar and consecutive binding events. Such binding events could take place at a platform, which offers multiple binding-sites, which can be regulated. In the transcription machinery there are at least three possible binding sites for consecutive binding events. These are the C-terminal domain (CTD) of Rpb1, the largest subunit of RNAPII, which carries 26 repeats in yeast, the C-terminal region (CTR) of the general elongation factor Spt5, which consists of 15 repeats in yeast, and the nascent mRNA itself. Since TREX has several subunits with RNA binding motifs and was shown to bind to RNA, we first inspected, whether efficient recruitment of TREX needs RNA at the site of transcription. In a study by (Abruzzi et al. 2004), Yra1 and Sub2 were shown to be at least partially sensitive to RNase digestion during a ChIP assay, but the Hpr1 (THO) signal was not sensitive to RNase digestion. However, the approach used in this study bears a technical shortcoming: The digestion of the RNA is carried out after the crosslinking. Thus, proteins which need RNA to be recruited or to be stabilized at the transcription site can be cross-linked to other proteins or the DNA by the formaldehyde and would therefore be in-sensitive to RNase digestion in the ChIP assay. To overcome this problem, we engineered a reporter construct, which can be integrated into the yeast genome: It is controlled by a galactose promoter to induce high expression levels for a good signal to noise ratio. A GFP coding sequence is followed by a sequence coding for a hepatitis δ ribozyme in the reporter (ribozyme sequence as in Fong et al. 2009) and integrated into *YCT1*, a non-essential yeast gene, in a way that the former start codon is positioned directly at the 3' end of the ribozyme sequence. As soon as the ribozyme is transcribed, it forms a hammerhead like structure and cleaves itself *in vivo* and releases the whole mRNA 5' of the ribozyme together with the RNA bound proteins. If a protein binds to the nascent RNA or needs the RNA to be recruited or stabilized at the transcription site, the occupancy of each protein should drop downstream of the cleavage site. For a more accurate analysis we used a second reporter construct, which served as a negative control: due to a single base substitution the ribozyme cleaves itself with app. 1000-fold slower kinetics (Fong et al. 2009) and can therefore be considered as inactive. The reporter is depicted in Figure 17.

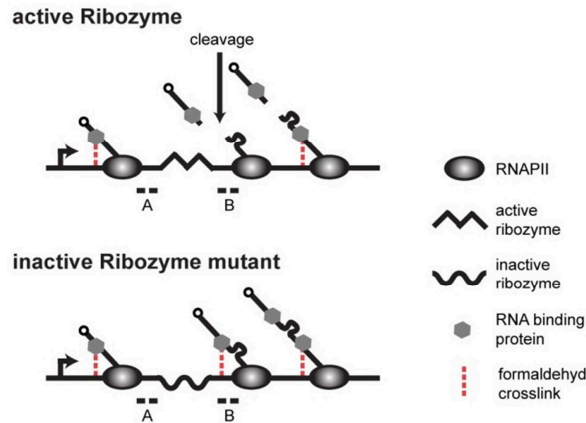


Figure 17: Ribozyme assay scheme.

Proteins tethered to chromatin by the mRNA will be lost by self-cleavage of the mRNA at the ribozyme sequence (upper panel). As negative control served a construct in which the ribozyme was inactivated by mutation of one base pair (lower panel). The occupancy after the cleavage (B) was normalized to the signal before the cleavage site (A). Further downstream, when the RNA is long enough, the cleavage event will not influence the occupancy of RNA-binding proteins any more (both panels, rightmost part of the pictures).

RNAPII occupancy was - as expected – not sensitive to RNA cleavage *in vivo* (Figure 18, Rpb3). The signal of all TREX subunits, Nab2 and Npl3 significantly decreased to about 70% app. 100 bp downstream of the cleavage site (Figure 18). Thus, all proteins assayed, which are involved in mRNA export, need intact nascent mRNA for their recruitment.

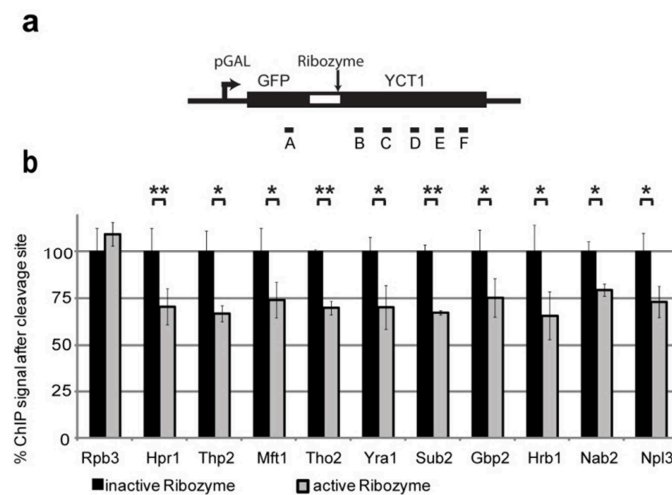


Figure 18: TREX, Nab2 and Npl3 depend on the nascent mRNA for efficient recruitment or occupancy.

(a) Scheme of the reporter containing the hepatitis virus ribozyme. A sequence coding for the GAL1 promoter, GFP and the hepatitis δ ribozyme (active or inactive mutant) was inserted 5' of the nonessential *YCT1* gene. (b) For each protein the ChIP signal after the cleavage site (B in a) was normalized to the signal before the cleavage site (A in a) and set to 100% for the inactive ribozyme (black bars). The ratio of B/A for the active ribozyme was calculated relative to the inactive ribozyme (grey bars). Whereas the signal for RNAPII (Rpb3) is unaffected by cleavage of the RNA, the signals for all TREX components, Nab2 and Npl3 dropped to about 70% indicating a at least partial RNA-dependent recruitment of these mRNA-binding proteins. Results of at least 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$).

Since all TREX subunits were RNA dependent in their recruitment, we asked, whether RNA can mediate long range interactions in a ChIP assay, as expected by a length dependent recruitment over several hundred bp as shown in Figure 13. To analyze this possibility, we determined the levels of the proteins at several positions downstream of the cleavage site. Indeed, while occupancy of Hpr1 and Sub2 was decreased 0.1 and 0.4 kb downstream of the cleavage site compared to the inactive ribozyme, the signals of the active and inactive ribozyme reporter were indistinguishable 0.7 kb 3' of the cleavage site (Figure 19). This indicates, that the RNA can only mediate effects in a ChIP (under the conditions used) for less than 0.7 kb and that the length dependency observed for THO, Sub2 and Yra1 is not mediated solely the RNA.

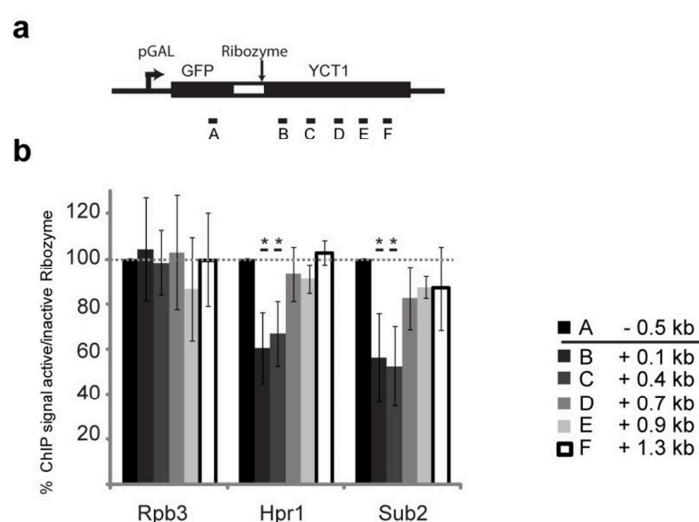


Figure 19: RNA dependent occupancy cannot be detected over more than 400-700 bp

ChIP signals for the TREX components Hpr1 and Sub2 and the RNAPII subunit Rpb3 before (A) and after (B-F) the ribozyme cleavage site were normalized to the signal in the inactive ribozyme reporter. Recruitment of RNAPII (Rpb3) was independent of RNA cleavage. Approximately 550 bp after the cleavage site the ChIP signals of Hpr1 and Sub2 became independent of the cleavage. Results of at least 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$).

5. CTD phosphorylation is essential for TREX recruitment *in vivo*

The second recruitment platform in transcription, which could mediate the length dependency, is the C-terminal domain of Rpb1, the largest subunit of RNAPII. The CTD is differentially phosphorylated during the transcription cycle and known to orchestrate the recruitment of a multitude of proteins involved in mRNA processing (reviewed in Zhang et al. 2012). The Tyr1 and Ser2 CTD phosphorylation have both been shown to occur throughout elongation phase of transcription on a genome-wide level (Mayer et al. 2010, Tietjen et al. 2010, Mayer et al. 2012a). Interestingly, both phosphorylations

showed a similar meta profile and length dependency as THO, Sub2 and Yra1 and also the Pearson correlation coefficient was very high between their recruitment profiles (Figure 20).

However, the Tyr1 phosphorylation signal dropped, similar as THO, Sub2 and Yra1 at *app*. the polyadenylation site, while the Ser2 phosphorylation signal persisted until the termination site (Figure 20), leading to the question whether one or both of the CTD phosphorylation marks are involved in recruiting TREX.

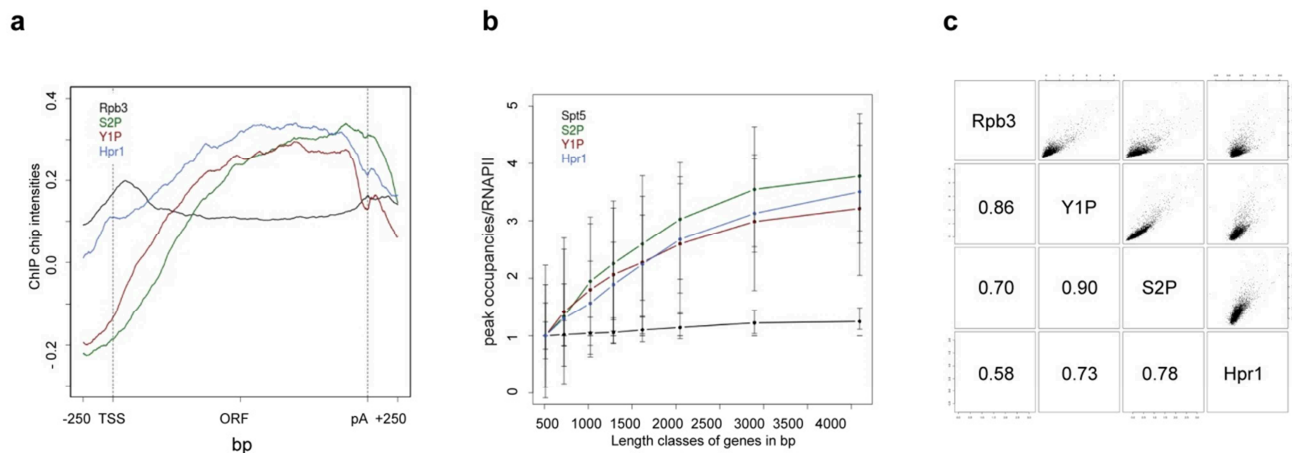


Figure 20: Tyr1 and Ser2 phosphorylations are length dependent and correlate with THO, Yra1 and Sub2.

(a) The meta gene occupancy profiles of S2 and Y1 CTD phosphorylation resembled the ones of TREX (Hpr1) **(b)** The peak occupancies of Hpr1, S2P and Y1P increased with gene length. Graph as in Figure 7. **(c)** Pearson correlation coefficients for the peak occupancies of Rpb3, Y1P, S2P and TREX (Hpr1). Hpr1 signal was better correlated with Y1P or S2P than with Rpb3.

We thus tested whether Tyr1 and/or Ser2 phosphorylation is needed for THO/TREX recruitment *in vivo*. Therefore TREX recruitment was assessed in Rpb1-CTD mutants. A strain with a Rpb1-CTD truncated to 14 *wt* repeats served as a *wt* control. The S2A mutant Rpb1-CTD carried 8 *wt* repeats and 7 repeats with Ser2 to Ala substitutions, which are under normal growth conditions (YPD, 30°C) viable (West and Corden 1995). We also constructed a similar mutant for Tyr1 with five *wt* and 9 Tyr1 to Phe substitution repeats. The occupancy levels of RNAPII (Rpb1) were largely unaffected in both mutants compared to the wildtype (Figure 21 a). Ser2P and Tyr1P levels in both mutants, the Ser2 Ala as well as the Tyr1 Phe, decreased as expected. Furthermore, the phosphorylation levels of Ser2 decreased in the Y1F mutant and the Tyr1 phosphorylation decreased in the S2A mutant, indicating a possible interdependency of Tyr1P and Ser2P (Figure 22, Figure 21). However, we cannot rule out that the efficiency of the immunoprecipitation was influenced by the small changes in the neighboring residue changing thereby the epitope of the antibodies slightly and hence lowering the ChIP signal. At least for the Tyr1P antibody, it was shown that the antibody recognizes Tyr1P and

Tyr1P-Ser2P repeats (Mayer et al. 2012a) *in vitro* equally well, indicating that at least in this case the epitope change from Ser to Ala at position 2 of the CTD might not be problematic.

Importantly, in both CTD mutants the levels of THO (Hpr1) were decreased (Figure 21 d). Furthermore, also the recruitment of the TREX subunits Yra1 and Sub2 was decreased in the Ser2 to Ala mutant (Figure 22 a, b), showing that proper CTD phosphorylation is needed to promote THO/TREX recruitment *in vivo*.

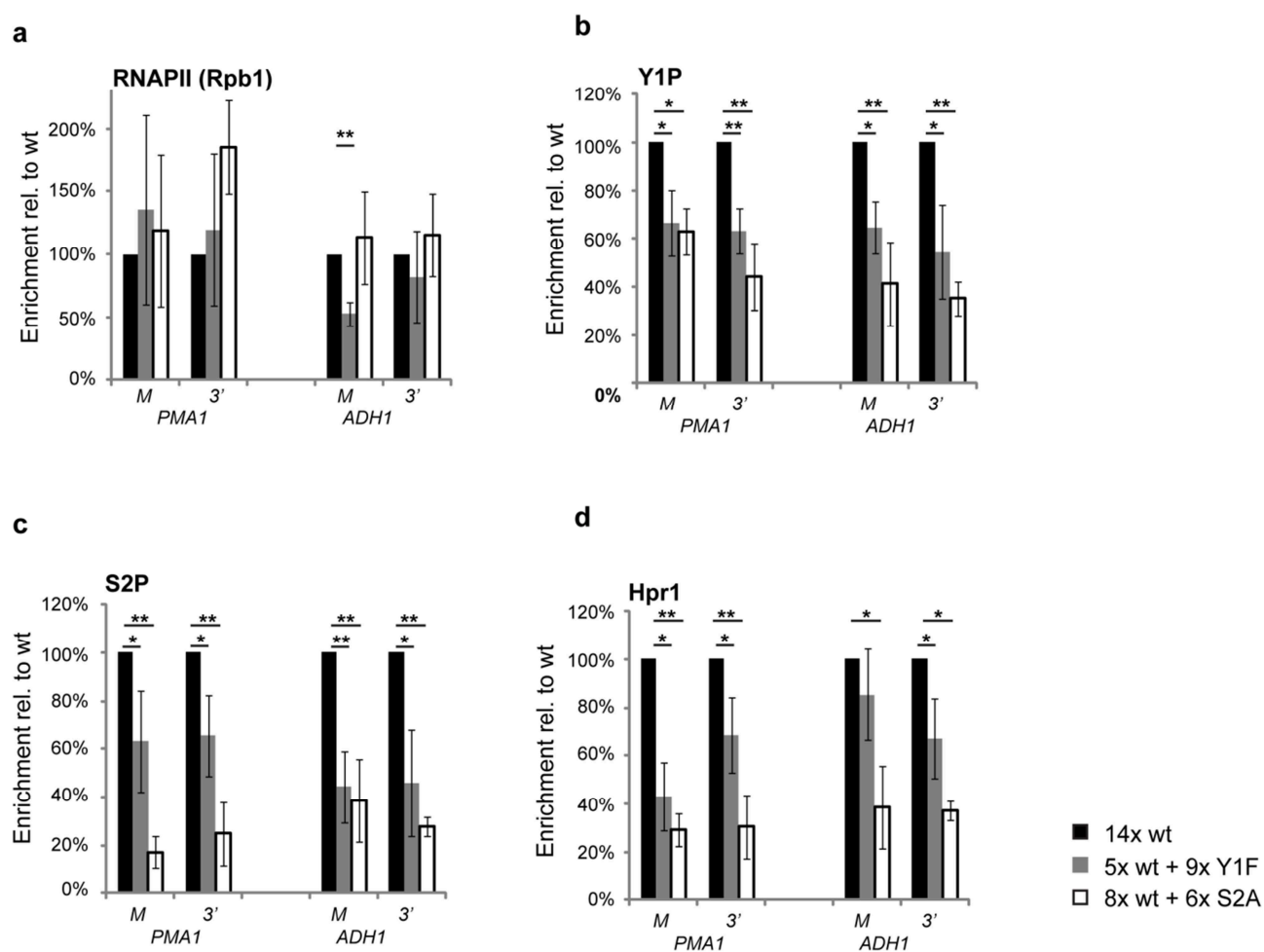


Figure 21: Proper Tyr1 and Ser2 CTD phosphorylation is essential for efficient THO recruitment.

The occupancies of (a) RNAPII (Rpb1), (b) Y1P, (c) Ser2P and (d) THO (Hpr1) in the Ser2 Ala (white bars) and the Tyr1 Phe (grey bars) mutant strains were calculated relative to the occupancy in a strain with 14 wild-type CTD repeats (black bars). Results of at least 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$). Primer positions are depicted in Figure 8 a.

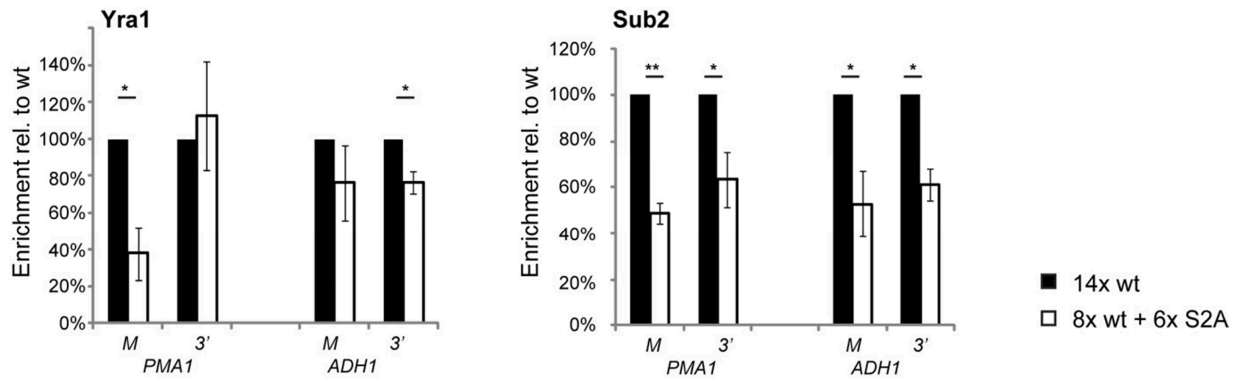


Figure 22: Proper Ser2 CTD phosphorylation is needed for efficient TREX recruitment.

The occupancies of the TREX subunits Yra1 and Sub2 in the Ser2 Ala (white bars) mutant strains was calculated relative to the occupancy in a strain with 14 wild-type CTD repeats (black bars). Results of at least 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$).

Consistently, analysis of the ChIP-chip data for non-coding RNAPII genes revealed, that Ser2P levels were much lower at sn/snoRNA genes than the Tyr1P levels and also THO, Sub2 and Yra1 levels stayed at low values at sn/snoRNA genes, so that it seemed more likely, that THO/TREX is recruited via Ser2P (Figure 23).

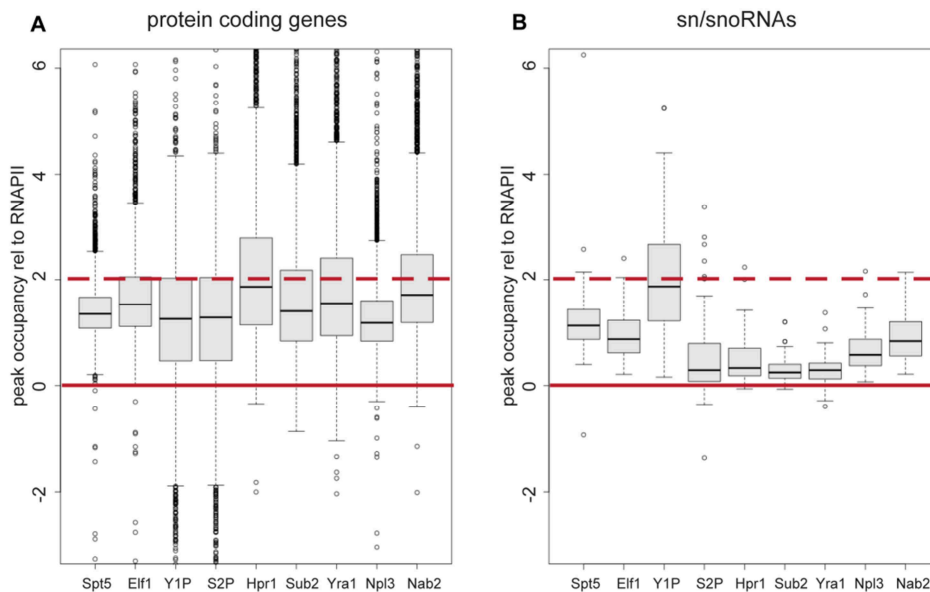


Figure 23: Ser2P levels are lower on sn/snoRNA genes, while Tyr1P is high

(a) Peak occupancies of the indicated proteins relative to RNAPII (Rpb3) on protein coding genes. The lower and upper borders of the boxes reflect the 25% and 75% quantiles, respectively, the black lines are the median values and the whiskers extend to the 1.5-fold inter quartile range. The red line gives the ratio 0, corresponding to no recruitment, and the dashed red line represents a ratio of 2. **(b)** Peak occupancies as in (a) but for sn/snoRNA genes.

Due to the assumed interdependency of Tyr1P and Ser2P it was not possible to unambiguously determine which one of the two phosphorylation signals is necessary for TREX recruitment *in vivo*, especially since further indirect effects by the mutations cannot be ruled out. However, proper CTD phosphorylation, most likely on Ser2 of the CTD, were clearly involved in recruiting TREX to the genes *in vivo*.

6. THO is recruited to the genes independently of Yra1

Yra1, a subunit of the TREX complex, was shown recently to bind to the Ser2-Ser5 diphosphorylated CTD *in vitro* and that the N-terminal 76 aminoacids of Yra1 are responsible for the interaction (MacKellar and Greenleaf 2011). Deletion of this region, termed phospho-CTD-interaction-domain (PCID), abrogated Yra1 recruitment to the transcription site *in vivo* (Figure 24 and MacKellar and Greenleaf 2011). Interestingly, the same domain was described to bind to Sub2 (Strässer and Hurt 2001) and Pcf11 (Johnson et al. 2009). Since the interaction of Yra1 with Pcf11 seems to open a Sub2 (TREX) independent pathway for recruiting Yra1 to genes (Johnson et al. 2009), it remained unclear whether the whole TREX complex is recruited to the genes via the PCID of Yra1, or whether the interaction of Yra1 with TREX, Pcf11 and the CTD are mutually exclusive to each other. To clarify if the whole TREX complex is recruited via the PCID of Yra1, we constructed a PCID deletion mutant of Yra1 lacking the N-terminal 76 amino acids, tagged with 3xHA (Figure 24 a) and assessed THO/TREX recruitment by ChIP. Although Yra1 recruitment levels vanished almost completely in the *yra1-ΔPCID* mutant, RNAPII (Rpb1) levels stayed almost unchanged (Figure 24 b). Importantly, THO subunits (Hpr1 and Mft1, Figure 24 c) were still recruited to wild type levels in the *yra1-ΔPCID*. Sub2 levels were partially reduced in the *yra1-ΔPCID* mutant to *app* 85% of the *wt* levels (Figure 24 c). This demonstrates that the Yra1 is not recruiting the rest of TREX complex to the site of transcription.

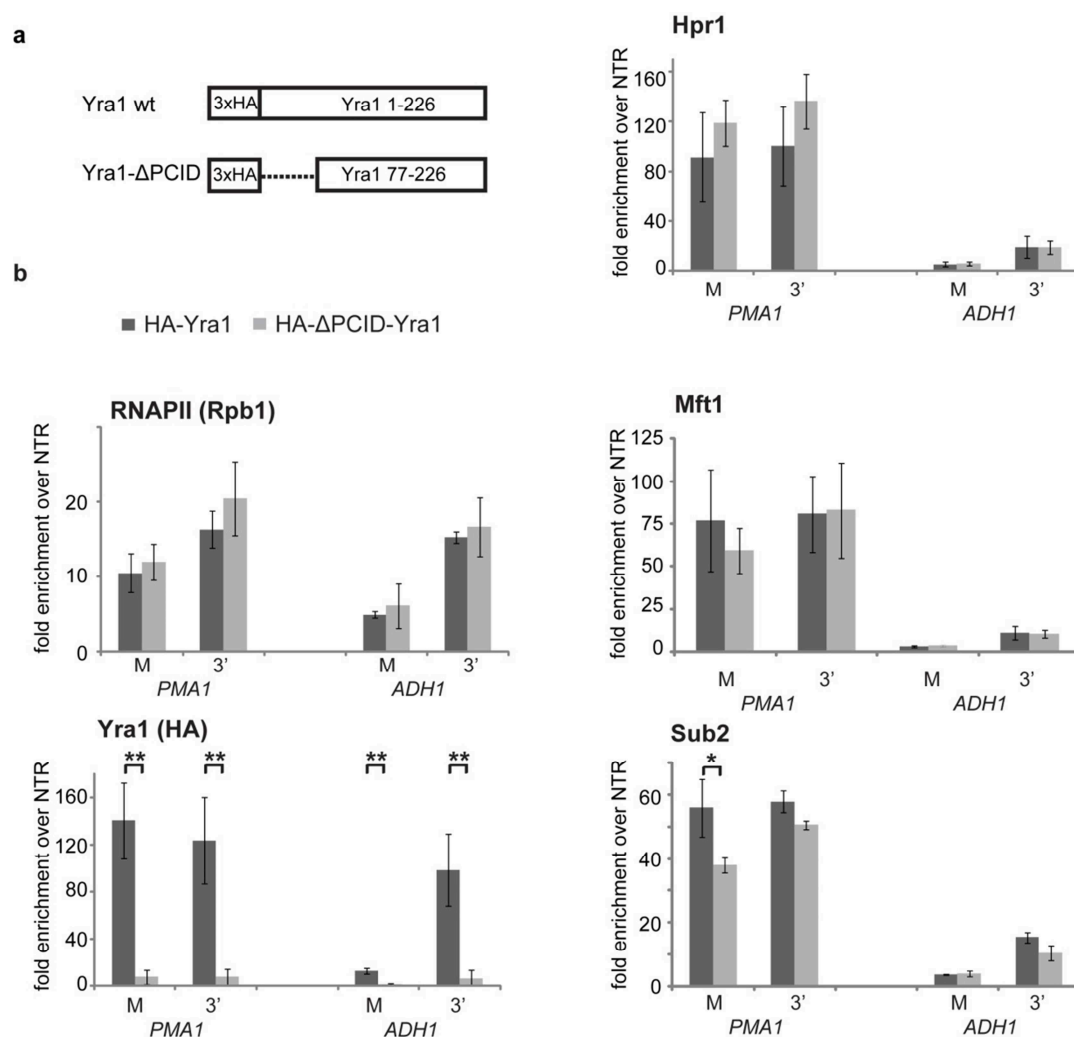


Figure 24: THO/TREX recruitment is independent of Yra1.

(a) Scheme of the Yra1 wild-type and the Δ PCID mutant, which was constructed as in (MacKellar and Greenleaf 2011). The N-terminal 76 amino acids were deleted in the Yra1- Δ PCID mutant. **(b)** Deletion of the PCID of Yra1 results in loss of Yra1 from the gene, but does not affect THO/TREX recruitment. The occupancies of RNAPII (Rpb1), Yra1, THO (Hpr1 and Mft1) and Sub2 (TREX) at *PMA1* and *ADH1* in *YRA1* wild-type (dark grey bars) and *yra1- Δ PCID* cells (light grey bars). Results of at least 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$).

Since THO/TREX was recruited independently of Yra1, the question arose, whether the TREX complex is still correctly assembled and if the PCID of Yra1 is needed for CTD interaction or for interaction with TREX. Interestingly, full-length Yra1 copurified with TREX (Hpr1-TAP, Figure 25) whereas Yra1- Δ PCID did not copurify. All other TREX subunits copurified with Hpr1-TAP in the *yra1- Δ PCID* mutant in similar levels as in the full-length *YRA1* strain, except for Sub2, which levels partially decreased. Since Sub2 and Yra1 were shown to interact directly (Strässer and Hurt 2001), Sub2 might be partially lost from the TREX complex due to a missing stabilizing interaction with Yra1 or lowered binding of Sub2 to the TREX complex in the *yra1- Δ PCID* mutant.

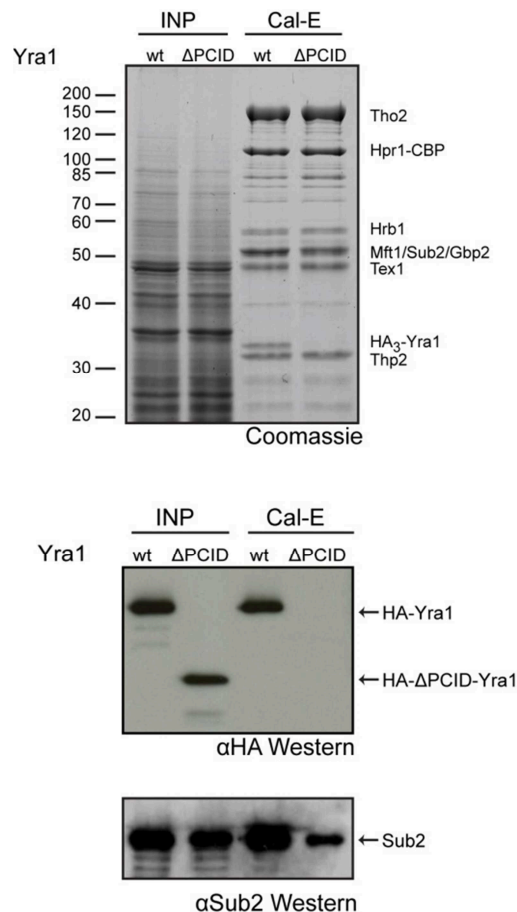


Figure 25: Yra1- Δ PCID does not assemble into the TREX complex

The PCID of Yra1 is essential for incorporation of Yra1 into TREX *in vivo*. TREX was purified from an *HPR1-TAP* strain expressing wild-type Yra1 (wt) or Δ PCID-Yra1 (Δ PCID). Lysates (INP) and eluates of the Calmodulin purification step (Cal-E) were stained with Coomassie (upper panel) and HA-tagged Yra1 was detected by Western blotting against HA (lower panel). Δ PCID-Yra1 does not copurify with TREX. Sub2 (lowest panel) does still copurify with TREX, but with lower efficiency.

The observed faulty incorporation of *yra1- Δ PCID* into the TREX complex might be due to three reasons: Firstly, the PCID might be necessary for the interaction of Yra1 with the other TREX subunits, which is in line with (Strässer and Hurt 2001), which showed, that the N-terminus is directly interacting with Sub2, which association with TREX is also decreased in the *yra1- Δ PCID* mutant. Secondly, the *yra1- Δ PCID* was shown to mislocalize to the cytoplasm (MacKellar and Greenleaf 2011) and might therefore not be able to interact with TREX and thirdly, the PCID deletion impairs interaction of Yra1 with Pcf11, which is needed for correct Yra1 recruitment (Johnson et al. 2009). In summary, Yra1 is not needed for recruitment of THO/TREX to the transcription machinery and the PCID of Yra1 most likely mediates the interaction of Yra1 with the whole TREX complex.

7. The THO complex binds directly to Ser2-Ser5 diphosphorylated CTD repeats

Since THO/TREX needs correct phosphorylation of the Rpb1-CTD for efficient recruitment *in vivo*, but Yra1 does not mediate the recruitment of TREX to the chromatin, the question arose, whether THO/TREX directly interacts with the CTD. To assess whether THO/TREX recruitment is mediated by direct interaction and which phosphorylations are needed, we performed *in vitro* pull-down experiments with highly purified THO complex. The THO complex was purified under high salt conditions yielding a pure THO complex comprised of Tho2, Hpr1, Mft1, Thp2 and Tex1. Importantly Yra1, as well as Sub2, Hrb1 and Gbp2 were quantitatively removed (Figure 26).

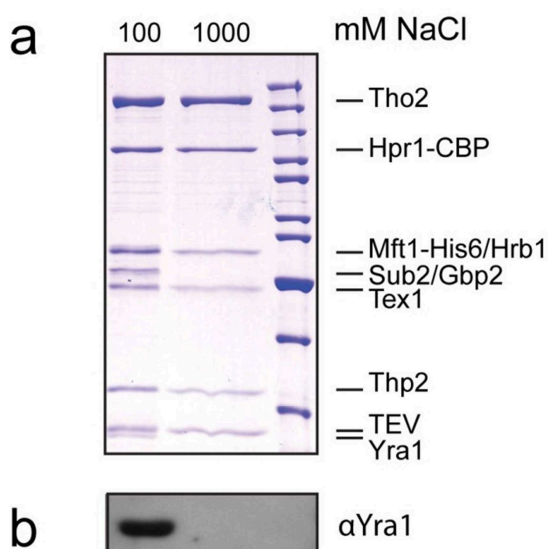


Figure 26: The THO complex can be purified without Yra1 using high salt conditions.

THO complex was purified from a strain expressing C-terminally TAP-tagged Hpr1 and C-terminally His₇-tagged Mft1 in two steps using IgG and NiNTA affinity purification under low salt (100 mM NaCl) and high salt (1000 mM NaCl) conditions, yielding the whole TREX complex or the THO complex, consisting of Tho2, Hpr1, Mft1, Thp2 and Tex1. **(a)** Coomassie staining of NiNTA affinity purification-eluates. The identity of each protein was verified by mass spectrometry and is indicated to the right. **(b)** Yra1 is absent from high salt purified THO complex. Western blot against Yra1 using antibodies directed against Yra1.

For the pull-down assay phosphorylated, N-terminal biotinylated peptides were immobilized on Streptavidin Dynabeads. Binding of THO was assayed to unphosphorylated (0), Tyr1 (Y1P), Ser2 (S2P), Ser5 (S5P) monophosphorylated and to Tyr1-Ser2, Tyr1-Ser5 and Ser2-Ser5 diphosphorylated peptides. Each peptide contained three CTD consensus repeats, in which the C-terminal two repeats were phosphorylated as indicated above (sequences are given in Table 7). The unrelated Rix1 complex, which is involved in processing of the ITS2 sequences from 35S pre-rRNA (Nissan et al.

2004, Krogan et al. 2004), served as negative control. The Pcf11 complex, which was shown to bind to Ser2 phosphorylated peptides (Licatalosi et al. 2002, Lunde et al. 2010), served as a positive control. The THO complex bound weakly to the Ser2 and Ser5 monophosphorylated CTD peptides, and more strongly to Ser2-Ser5 diphosphorylated peptides, but not to Tyr1 phosphorylated repeats (Figure 27). Importantly, treatment of the peptides with alkaline phosphatase (AP) completely abolished CTD binding of THO, indicating that binding is in fact due to phosphorylation. This result is consistent with the requirement for proper Ser2 phosphorylation for efficient THO/TREX recruitment *in vivo* (Figure 21 and Figure 22).

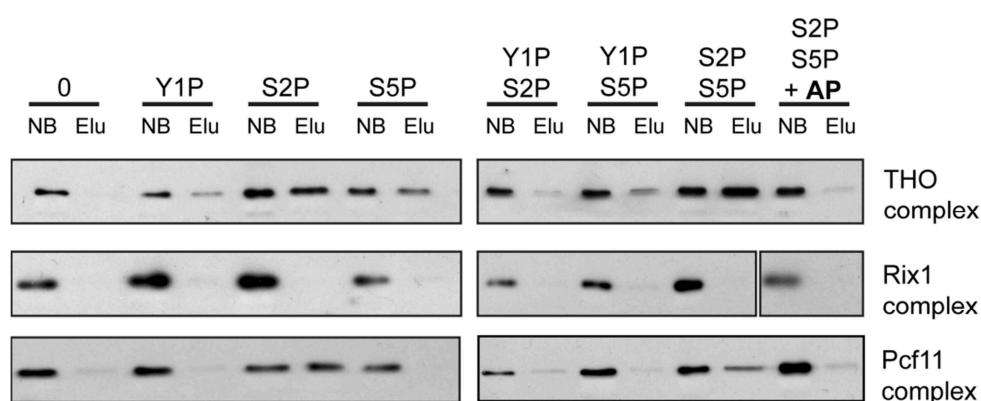


Figure 27: THO binds directly to Ser2-Ser5-diphosphorylated CTD-peptides

Pull-down experiments were performed with CTD peptides that were not phosphorylated (0), mono-phosphorylated on tyrosine 1 (Y1P), serine 2 (S2P) or serine 5 (S5P) or diphosphorylated on tyrosine 1 and serine 2 (Y1PS2P), tyrosine 1 and serine 5 (Y1PS5P) or serine 2 and serine 5 (S2PS5P) immobilized to Streptavidin Dynabeads. The THO complex binds to CTD peptides phosphorylated on serine 2 (S2P) and serine 5 (S5P) and more strongly to the serine 2 and serine 5 diphosphorylated CTD (S2PS5P). Binding of THO to the CTD is dependent on serine 2 and serine 5 phosphorylation since treatment with alkaline phosphatase (AP) of the serine 2 and serine 5 diphosphorylated CTD peptide abrogates binding of THO (S2PS5P + AP). The unrelated Rix1 complex served as negative control. Pcf11 was used as a positive control and binds – as expected – to the serine 2 monophosphorylated and serine 2 and serine 5 diphosphorylated CTD peptides. The TAP-tagged protein of each complex was detected by Western blotting against CBP (Hpr1 for THO, Rix1 for the Rix1 complex and Pcf11 for the Pcf11 complex). A representative experiment is shown.

The increase of the Ser2 phosphorylation from 5' to 3' over the gene in combination with the basal level of Ser5 phosphorylation throughout the ORF (Kim et al. 2009) therefore most likely explains together with the binding of THO to Ser2-Ser5 diphosphorylated repeats the increase in THO occupancy during transcription elongation and elucidates the THO/TREX recruitment mechanism on a molecular level: THO/TREX is most likely recruited to the transcription site by Ser2-Ser5 diphosphorylated CTD repeats.

8. C-terminal TAP-tag fusion to Tho2 impairs TREX recruitment

In the initial ChIP-chip experiments, we found that Tho2-TAP was not recruited properly to the genes. Tho2-TAP ChIP signals did not increase length dependent from 5' to 3' as observed for the other THO subunits, but showed only a non-length dependent basal occupancy throughout the whole ORF of the genes. (Figure 28 a, b). In contrast, TAP-Tho2 showed in the ChIP-chip the same length dependency and a similar meta profile as Hpr1 or Mft1 (Figure 28 a and b) and also had a high Pearson correlation coefficient to the other THO subunits (Figure 15). To determine the occupancy of Tho2-TAP more quantitatively we performed small scale ChIP analysis of Tho2-TAP in comparison to TAP-Tho2. Interestingly, Tho2-TAP and TAP-Tho2 were both recruited to similar levels at the 5' of the gene, but TAP-Tho2 levels increased to the 3' (Figure 28 c).

We next asked whether the whole TREX complex recruitment is impaired and if the *THO2-TAP* strain might be a functional mutant for non-length dependent TREX recruitment. Therefore, we inspected THO/TREX recruitment in the *THO2-TAP* mutant. Since the protein A moiety of the TAP-tag interfered with the use of any other antibody in a ChIP assay, we made use of the Avi-tag (van Werven and Timmers 2006) to inspect the other TREX subunits Hpr1, Sub2 and Yra1. The Avi-tag, a bacterial biotin-acceptor peptide, is specifically biotinylated by the biotin ligase BirA, which was expressed in the assay from a plasmid. BirA was fused to a nuclear localization sequence for higher biotinylation efficiency (van Werven and Timmers 2006). The biotinylated proteins were purified via binding to Streptavidin beads.

To exclude that the decreased recruitment is due to changed transcription of the gene, we assayed the RNAPII levels by ChIP of Rpb1-Avi in the *THO2-TAP* and in the *wt* strain. RNAPII (Rpb1) levels did not change significantly (Figure 28 g), indicating similar transcriptional activity in the *wt* and *THO2-TAP* strain. Hpr1, Sub2 and Yra1 were also recruited to the 5' of the gene in similar levels in both the *wt* and the *THO2-TAP* strain. But just as Tho2-TAP, also Hpr1, Sub2 and Yra1 were impaired in their recruitment in the *THO2-TAP* strain throughout the ORF and only reach a certain basal level (Figure 28 c-f).

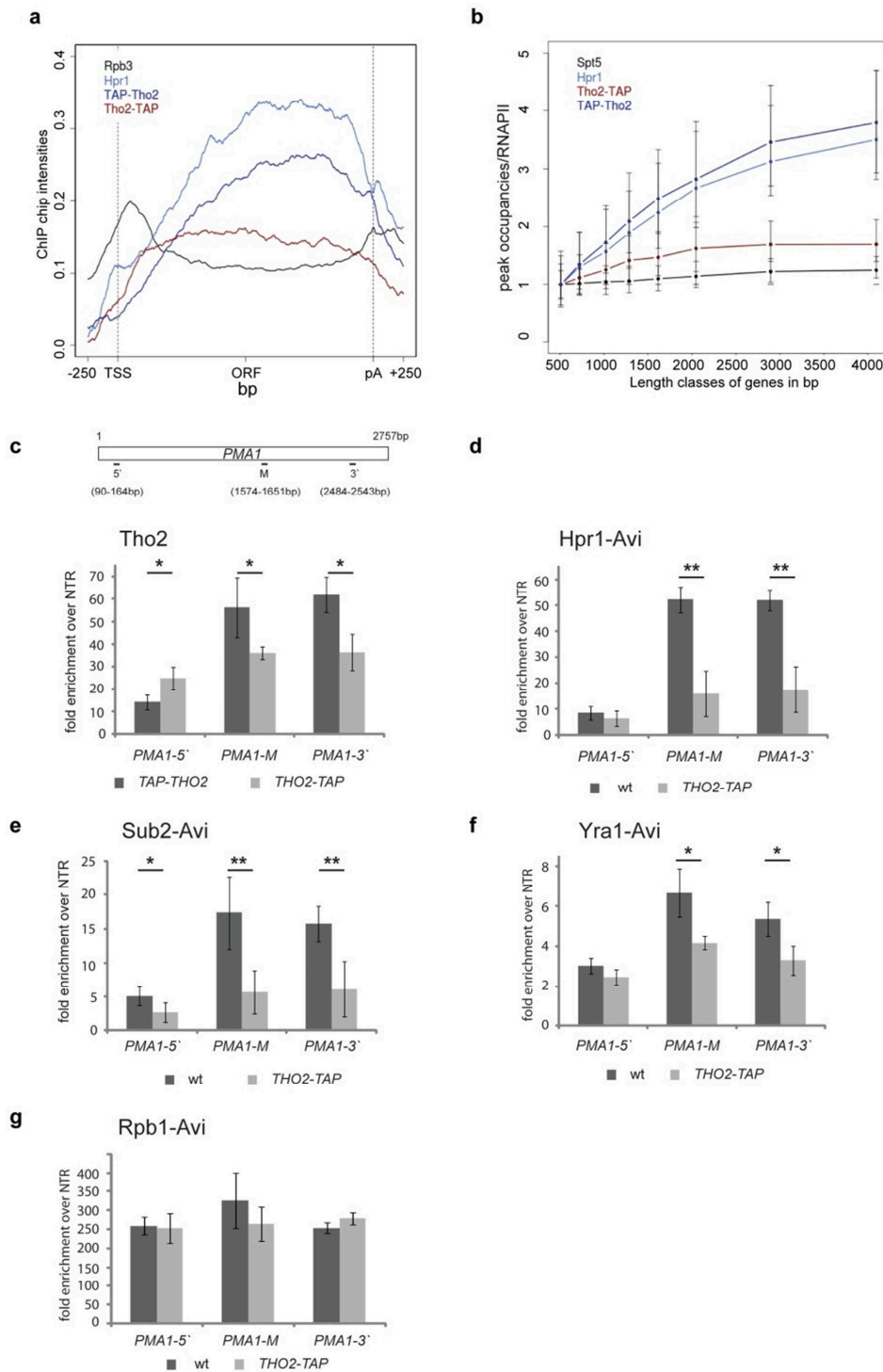


Figure 28: *THO2-TAP* impairs recruitment of *THO/TREX*

(a) Meta gene occupancy profiles of RNAPII (Rpb3), Hpr1, TAP-Tho2 and Tho2-TAP. TAP-Tho2 showed the 5'-3' increase typical for THO/TREX components, while Tho2-TAP was still recruited to genes, but did not increase throughout the ORF. The Y-intercept of Tho2-TAP was adjusted with -0.05 to superimpose Tho2-TAP and TAP-Tho2 at the transcription site for better visualization. **(b)** Occupancy of Tho2-TAP did not increase with gene length as the other TREX components. Peak occupancy of Hpr1, TAP-Tho2 and Tho2-TAP in comparison to the bona fide transcription elongation factor Spt5. **(c)-(f)** TREX was recruited to the PMA1 gene but its occupancy did not increase in the *THO2-TAP* mutant. Occupancies of Tho2 **(c)**, Hpr1 **(d)**, Sub2 **(e)** and Yra1 **(f)** at the PMA1 gene are changed in the *THO2-TAP* strain relative to the *TAP-THO2/wt* strain. **(g)** RNAPII (Rpb1) recruitment levels were still in the *THO2-TAP* strain as in *wt* cells. Results of 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$). To assess the occupancy of Hpr1, Yra1 and Sub2 in the presence of the TAP-Tag at Tho2, the other proteins were tagged with the Avi-tag.

Since defective assembly of the TREX complex could be the cause for impaired recruitment of the whole complex, the TREX complex assembly was assessed by purification of the complex from the *THO2-TAP* and *TAP-THO2* strain. Purifications proofed that the TREX complex is still correctly assembled and that all TREX complex subunits can be found in similar stoichiometric amounts in both strains (Figure 29).

In summary, the whole TREX complex is impaired in its recruitment in the *THO2-TAP* strain. The THO/TREX complex was still correctly assembled. Furthermore, TREX is not recruited length dependently in the *THO2-TAP* strain, giving the opportunity to analyze the function of the length dependency for the gene expression in the cell.

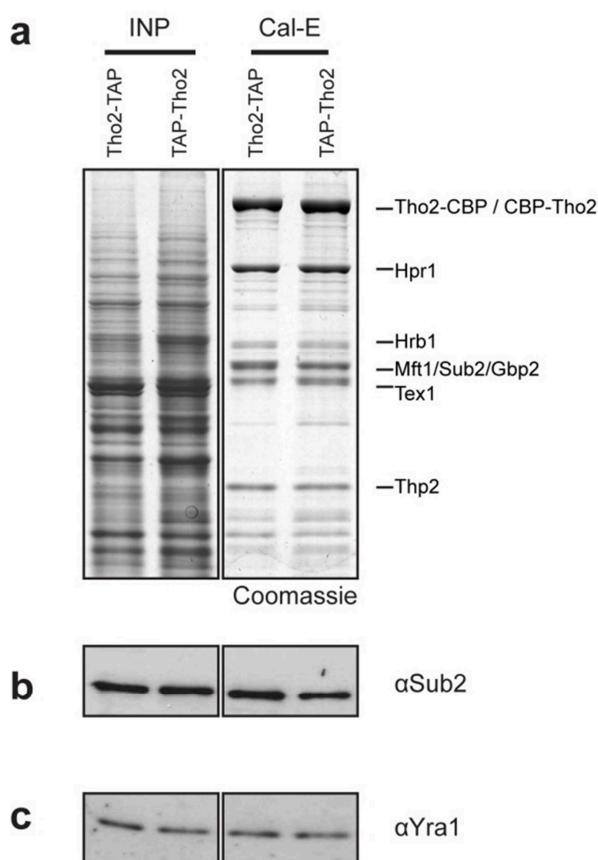


Figure 29: Tho2-TAP does not impair TREX assembly.

TREX complex was purified using a strain containing a C-terminal or an N-terminal TAP-tagged Tho2 under low salt conditions until the Calmodulin step. **(a)** Coomassie stain of the lysates (INP) and the eluates from the Calmodulin purification step (Cal-E). Identity of proteins was verified using Mass-spectrometry. Western blots using antibodies specific to **(b)** Sub2 and **(c)** Yra1 indicating, that Yra1 and Sub2 are both still efficiently assembled into the TREX complex of the *TAP-THO2* and *THO2-TAP* strains.

9. The 5' to 3' increasing recruitment of THO/TREX is important for correct gene expression

To assess the physiological relevance of the length dependent increase of THO/TREX, we analyzed the transcriptome of the *THO2-TAP* strain relative to the corresponding *wt* strain (in collaboration with Eoghan O'Dubhir and Frank Holstege). Interestingly, we found that down-regulated genes were on average significant longer than the average yeast transcript. Up-regulated genes were on average much shorter than the average transcript. In contrast, all other features analyzed were not enriched in the up- or down-regulated genes, like *e.g.* expression strength, GC content, convergent or divergent, or positioning of the +1 nucleosome or the type of promoter (Figure 30). Therefore, the length dependent recruitment of the THO/TREX complex is needed for correct expression of long transcripts in yeast.

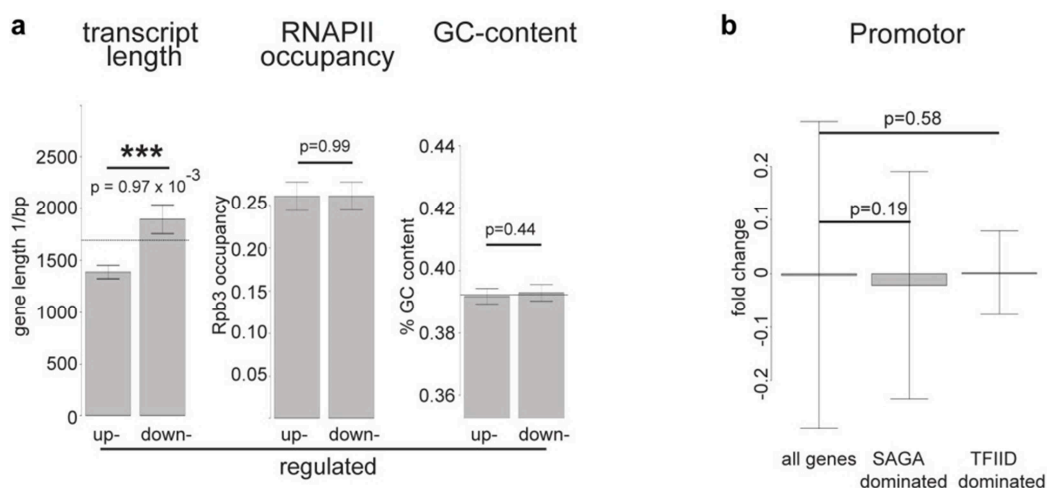


Figure 30: *THO2-TAP* impairs correct expression of long genes.

RNA of the *THO2-TAP* strain was isolated and hybridized to expression arrays relative to the corresponding wildtype. **(a)** Up and down regulated genes were analyzed for their average transcript length, RNAPII occupancy and GC content. Down-regulated genes were in average longer and up-regulated genes were in average shorter as the average yeast genes. No effects were observed for the transcriptional frequency, reflected by RNAPII occupancy or the GC content of the transcript. **(b)** Influence of the promoter on the expression changes were analyzed: SAGA and TFIID dominated promoters are not significantly different from all other genes in the *THO2-TAP* mutant.

10. The PCID of Yra1 genetically interacts with *THO2-TAP*

To determine the impact of the *yra1-ΔPCID* mutation on the recruitment of THO/TREX in the *THO2-TAP* strain, we generated double mutants. Interestingly, *yra1-ΔPCID THO2-TAP* strains were synthetic lethal and therefore both mutations interact genetically (Figure 31). This demonstrates the importance of the increasing recruitment of THO/TREX throughout the gene, because a second impairment of the mRNA export machinery (recruitment of Yra1) leads to cell death.

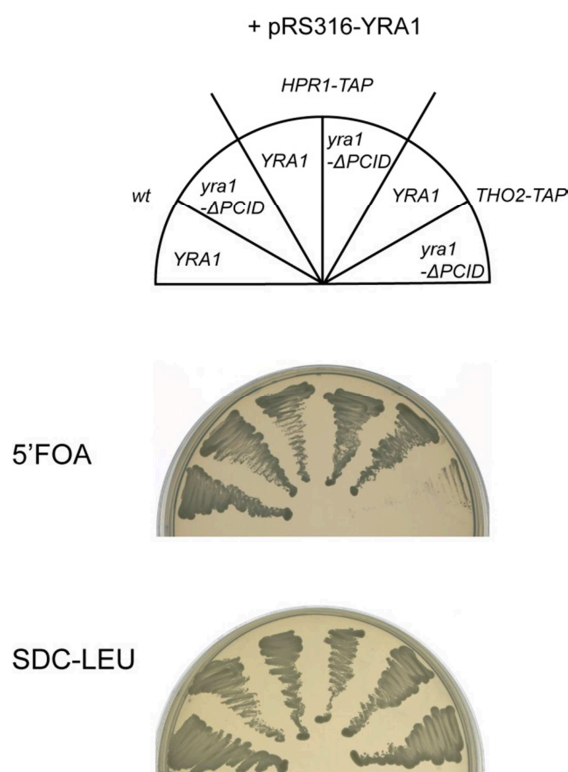


Figure 31: *THO2-TAP* is synthetic lethal with *yra1-ΔPCID*

Growth of strains expressing Yra1 or *yra1-ΔPCID* and either no tagged protein, Hpr1-TAP or Tho2-TAP and carrying the plasmid *pRS316-YRA1* on SDC(-LEU) and 5-FOA, which counter-selects against the Ura3-encoding *pRS316* plasmid. *yra1-ΔPCID* was synthetically lethal with *THO2-TAP*, which causes an aberrant TREX occupancy profile, but not with *HPR1-TAP*.

11. Tho1 recruitment recapitulates THO/TREX recruitment

Tho1, an hnRNP protein, is implicated in mRNA export and it was previously shown that Tho1 is co-transcriptionally recruited to the active chromatin (Jimeno et al. 2006). Since the Tho1 homolog CIP29 associates ATP dependently with the human TREX complex (Dufu et al. 2010), we aimed to analyze Tho1 recruitment to learn more about Tho1 function. Therefore Tho1-TAP genome-wide recruitment was assessed by ChIP-chip. Tho1 was recruited similar to THO/TREX during transcription elongation and left the site of transcription at *app.* the polyadenylation and cleavage site (Figure 32). Furthermore, Tho1 was also recruited length dependently, just as THO/TREX. This was also reflected by a high Pearson correlation coefficient between THO/TREX and Tho1 (Figure 33). Interestingly, the Pearson correlation coefficient between Tho1 and Hpr1/Yra1 was even higher than between Hpr1 and Yra1 (Figure 33), indicating a strong relationship in the recruitment of Tho1 and THO/TREX. However, Tho1 correlates only very weakly with RNAPII, indicating that Tho1 recruitment might not be only regulated strictly by occupancy of RNAPII at the gene but also by other factors, *e.g.* by occupancy of THO/TREX (Figure 33).

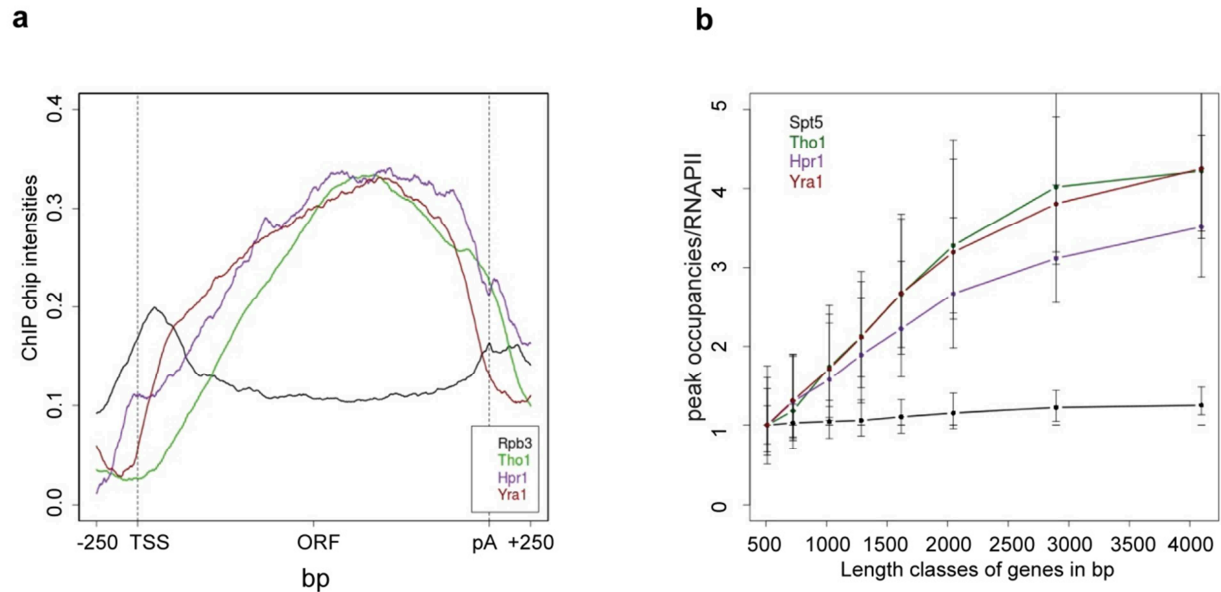


Figure 32: Tho1 is recruited during elongation phase and Tho1 recruitment is length dependent.

Meta gene occupancy profiles of RNAPII (Rpb3), Tho1, Hpr1 (THO) and Yra1 (TREX). Similar as the THO/TREX subunits, Tho1 occupancy signal increased throughout transcription elongation and dropped at around the polyadenylation site. **b.** Tho1 peak occupancies increased with gene length, similar to the THO/TREX subunits Hpr1 and Yra1. Genes were subdivided according to the indicated length classes and their occupancy was normalized for transcriptional frequency (Rpb3 ChIP signal). Length classes are: A (512-723 bp), B (724-1023 bp), C (1024-1286 bp), D (1287-1617 bp), E (1618-2047 bp), F (2048-2895 bp), G (2896-4095 bp) and H (4096-5793 bp).

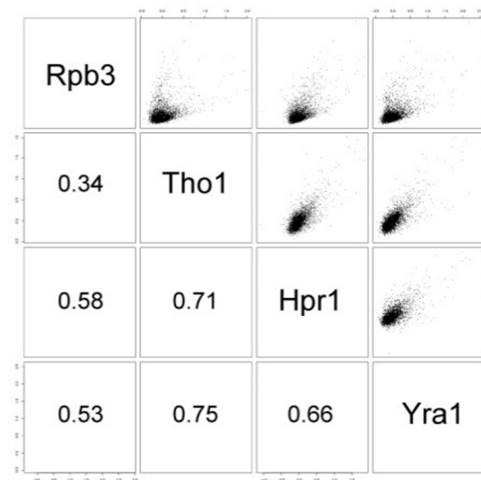


Figure 33: Tho1 occupancy correlates very well with THO/TREX

For each protein the peak occupancies for each gene (90th percentile of each profile for each gene) were calculated. The numbers give the Pearson correlation coefficient of two proteins, in the graph the black points represent the peak occupancies. Tho1 Pearson correlation with the THO/TREX subunits Hpr1 and Yra1 was very high. It was even higher between Yra1 and Tho1 than between Hpr1 and Yra1. Interestingly, Tho1 displayed a very weak correlation coefficient to Rpb3.

Since the genome-wide occupancy of Tho1 was very similar to the recruitment of THO/TREX, Tho1 recruitment was also inspected for its RNA dependency. Indeed, the ribozyme assay revealed a

similar reduction of Tho1 signal after cleavage by the ribozyme, indicating that efficient Tho1 recruitment depends upon the nascent RNA (Figure 34). In summary, the sensitivity to RNA, the similar meta profiles of the occupancies, the length-dependency which was by now only shown for THO/TREX and the high correlations between the recruitment profiles of Tho1 and TREX indicated, that there might be a similar recruitment pathway and/or that Tho1 and TREX form together a functional entity.

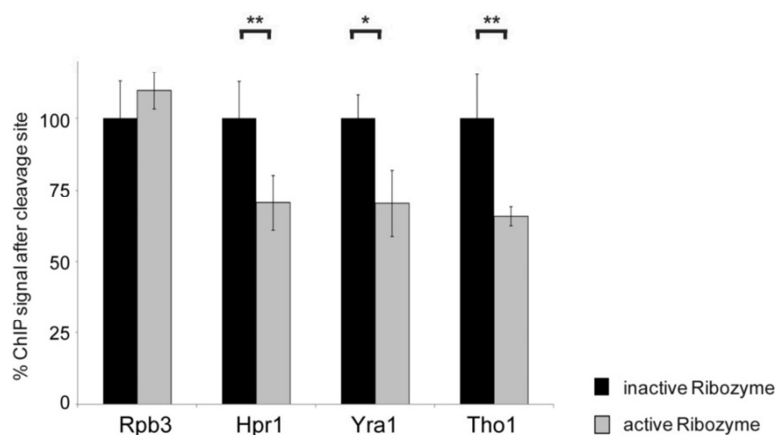


Figure 34: Tho1 recruitment to the transcription site is promoted by the nascent RNA

Ribozyme assay was performed and analyzed as described above (Figure 18). The signal of the active ribozyme was normalized by the signal of the inactive ribozyme, which was set to 100% (black bars). Whereas the signal for RNAPII (Rpb3) is unaffected by cleavage of the RNA, the signals for the TREX components, Hpr1 and Yra1 drop to about 70%. Tho1 also decreases to app. 70% showing similar dependency on RNA as THO/TREX. Results of at least 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$).

Next we assessed whether Tho1 recruitment is THO/TREX dependent, since Tho1 recruitment features recapitulated the ones of THO/TREX. Therefore, we analyzed Tho1 occupancy by ChIP in a $\Delta hpr1$ strain. As expected by a role of the THO complex in transcription elongation, RNAPII occupancy (Rpb3) was lower (app. 50%) in the $\Delta hpr1$ strain as in the *wt* strain (Figure 35 a). However, Tho1 levels were reduced to app. 10% in the ORF of *PMA1* and *YEF3* (Figure 35 b), reflecting almost background levels of the ChIP experiment and showing, that indeed THO/TREX is needed for Tho1 recruitment.

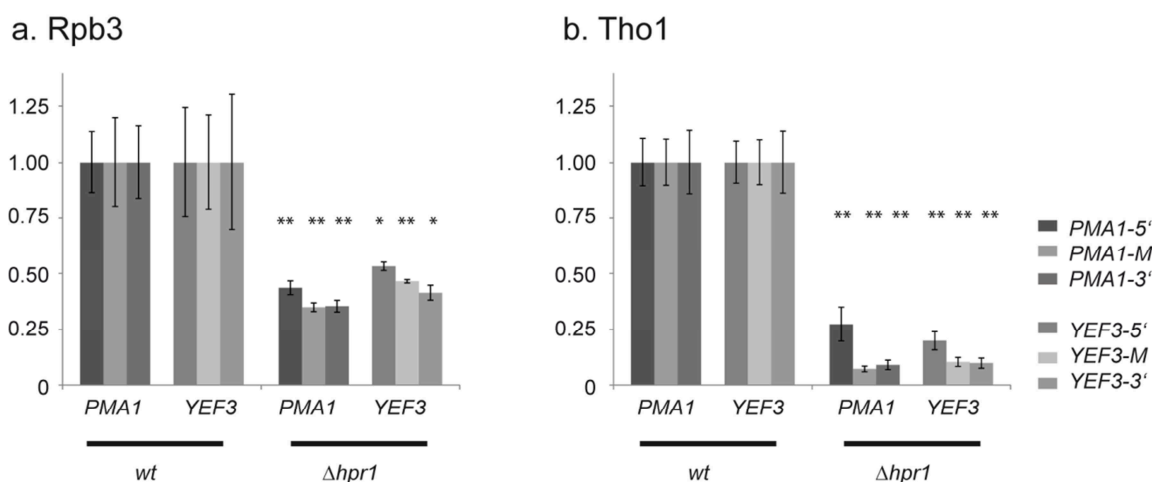


Figure 35: $\Delta hpr1$ abolishes Tho1 recruitment

(a) RNAPII (Rpb3) occupancy on the genes *PMA1* and *YEF3* was reduced to app. 50% in a $\Delta hpr1$ strain. (b) Tho1 levels were reduced to less than 10 % in the ORF of *PMA1* and *YEF3* upon $\Delta hpr1$, reflecting almost background levels. ChIP enrichments over NTR were normalized by the wildtype level of each gene. Results of 3 independent experiments are shown (mean \pm SD; **: $p < 0.01$; *: $p < 0.05$).

12. The Spt5 CTR regulates recruitment of mRNA export factors

In addition to the RNA and the CTD of Rpb1, a third recruitment platform was described for transcription: The C-terminal region (CTR) of the general *bona fide* elongation factor Spt5. The CTR of Spt5 consists of 15 hexa-repeats in yeast, which can be phosphorylated during transcription elongation by Bur1 (Liu et al. 2009) and which was shown to recruit factors involved in mRNP biogenesis, like *e.g.* the cleavage factor CF1 (Mayer et al. 2012b) and the Paf1 complex (Qiu et al. 2012). To test whether correct phosphorylation of the Spt5-CTR is necessary for correct THO/TREX recruitment, we used a Spt5 mutant, in which the CTR repeats were mutated from Ser1 to Ala, to abolish CTR phosphorylation. RNAPII levels showed no significant changes in the S1A mutant compared to the wt (Figure 36 a) and also Rpb1-CTD Ser2 and Tyr1 phosphorylation were almost unchanged (Figure 36 b and c). Furthermore, the Spt5-S1A mutant was recruited to similar levels as the wt protein (Figure 36 d). As shown previously, Paf1 levels decreased significantly in the Ser1 Ala mutant (Figure 36 e) (Qiu et al. 2012). Unexpectedly, THO (Hpr1) occupancy was increased in the non-phosphorylated mutant compared to the wt, especially towards the 3' end of the gene (Figure 36 g). Furthermore, the levels of the TREX subunits Sub2 and Yra1 were increased in the unphosphorylated mutant (Figure 36 h and i). Interestingly, Tho1 levels, which was by now only observed to behave like the THO/TREX complex, significantly decreased throughout the gene (Figure 36 f). This could indicate a positive role of Spt5-pCTR in recruiting Tho1 and an inhibitory function for the recruitment of THO/TREX.

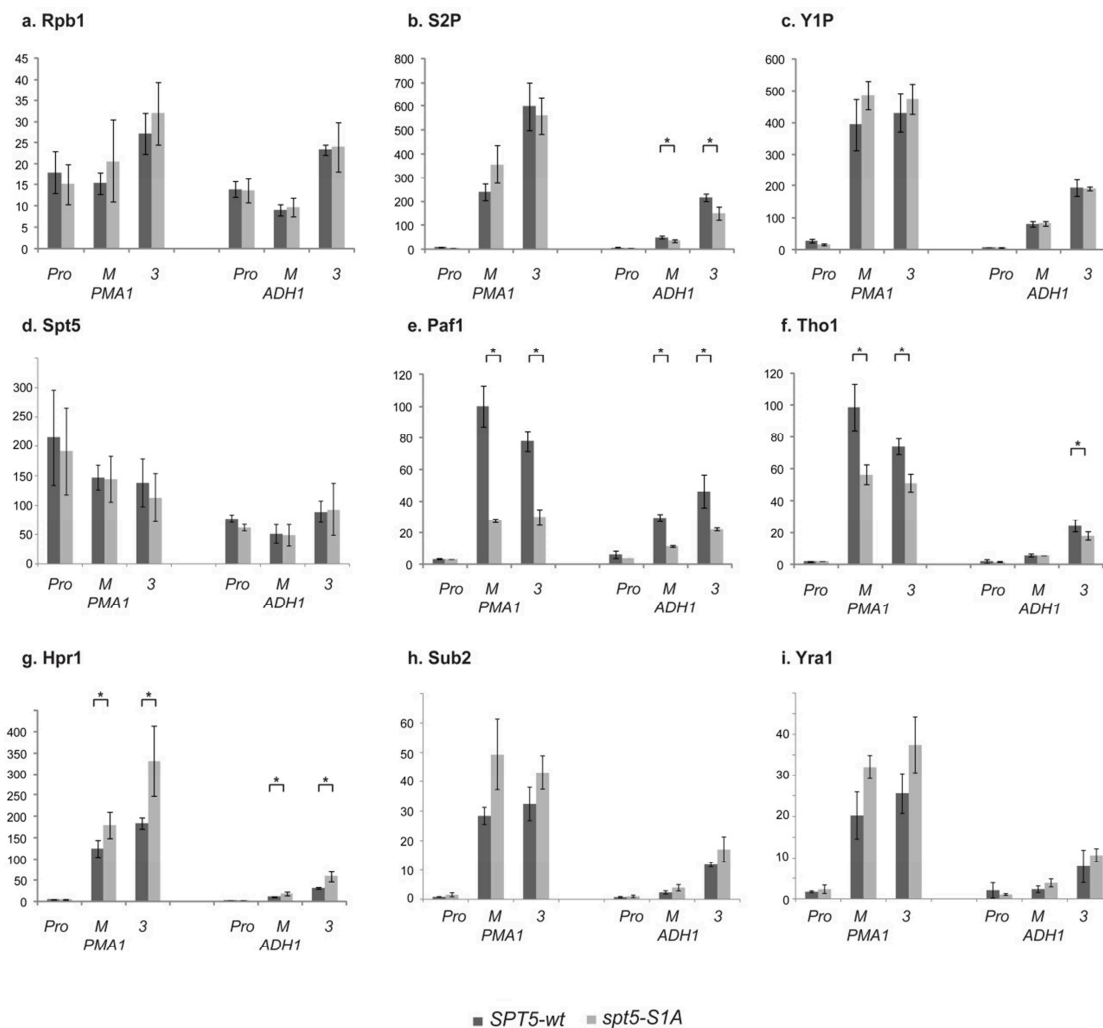


Figure 36: *Spt5*-CTR S1A mutations lead to increased THO/TREX and decreases *Tho1* occupancies

The fold enrichment over NTR of RNAPII (Rpb1) (a), Ser2 (b) and Tyr1 (c) phosphorylation, Spt5 (d), Paf1 (e), Tho1 (f) and the TREX subunits Hpr1 (g) Sub2 (h) and Yra1 (i) in the *Spt5*-CTR S1A (light grey bars) mutant strains and the corresponding wildtype *Spt5*-CTR S1S strain (dark grey bars). Results of at least 3 independent experiments are shown (mean \pm SD; *: p < 0.05). Primer positions are depicted in Figure 8 a.

To test whether the phosphorylation is a signal for recruitment or just the mutation itself abolished the recruitment, RNAPII, THO/TREX and Tho1 levels were assessed in a phospho-mimicry mutant, in which all 15 Ser1 of the *Spt5*-CTR repeats were substituted by an Asp residue (S1D). Indeed, while RNAPII levels and Ser2 phosphorylation levels were almost unchanged (Figure 37 A and B) in the phospho-mimicry mutant, THO/TREX levels (Hpr1, Sub2 and Yra1, Figure 37 D, E and F) significantly decreased in the phospho-mimicry mutant, indicating a negative regulation of TREX recruitment by phosphorylated *Spt5*-CTR. Interestingly, Tho1 levels almost recovered to more than 80% (Figure 37 C), similar as it has been observed for Paf1 complex, which was shown to directly bind to the phospho CTR of *Spt5* (Qiu et al. 2012a).

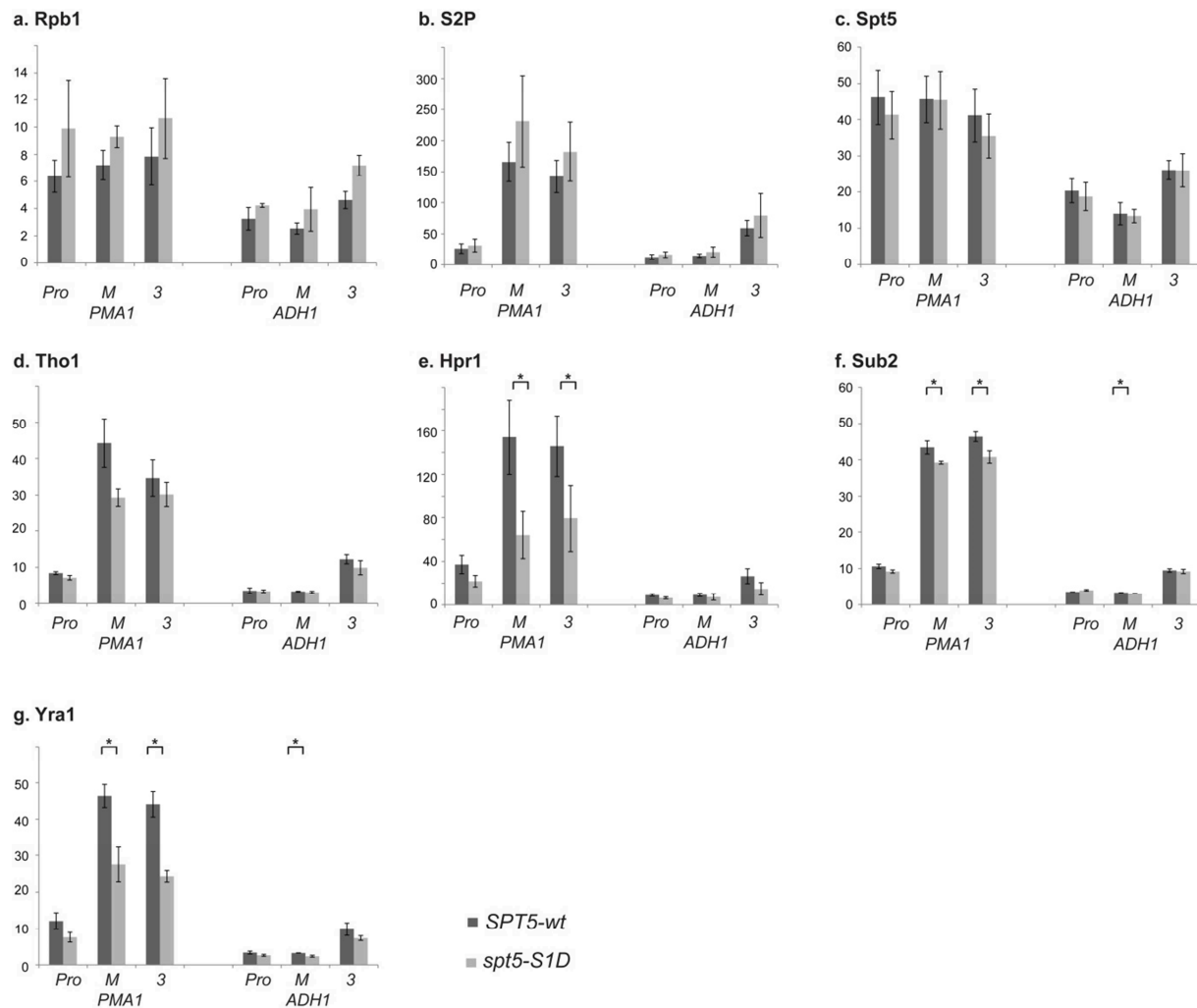


Figure 37: Spt5-CTR S1D phospho-mimicry mutants restore Tho1 levels and decrease THO/TREX levels

The fold enrichment over NTR of RNAPII (Rpb1) (a), Ser2 phosphorylation (b), Spt5 (c) Tho1 (d) and the TREX subunits Hpr1 (e) Sub2 (f) and Yra1(g)) in the Spt5-CTR S1D (light grey bars) mutant strains and the corresponding wildtype Spt5-CTR S1S strain (dark grey bars). Results of at least 3 independent experiments are shown (mean \pm SD; *: $p < 0.05$). Primer positions are depicted in Figure 8 a.

13. Tho1 overexpression decreases THO/TREX recruitment

Since Tho1 recruitment and THO/TREX recruitment showed opposing effects in the Spt5-CTR-S1A mutant, the question arose, whether Tho1 might control THO/TREX recruitment by negative regulation. To address this question, THO/TREX recruitment was assessed by ChIP in cells expressing genomically 6xHA-tagged Tho1 under the control of the endogenous promoter as well as carrying an empty plasmid (+ *pRS426*) and in cells additionally overexpressing Tho1-6xHA from a multicopy plasmid (*pRS426-THO1-6xHA*) to the genomically 6xHA tagged *THO1* (Figure 38 a). As hypothesized by the opposing effects in the CTR mutant, while RNAPII levels stayed at wild type levels at *YEF3* and *PMA1* (Figure 38 a). THO (Hpr1) (Figure 38 b) and Sub2 levels were significantly decreased in the overexpression strain relative to the *wildtype* strain. Interestingly, Yra1 levels did not change in the

THO1 overexpression strain. This indicates that Tho1 might indeed regulate the occupancy of THO and Sub2, but not Yra1.

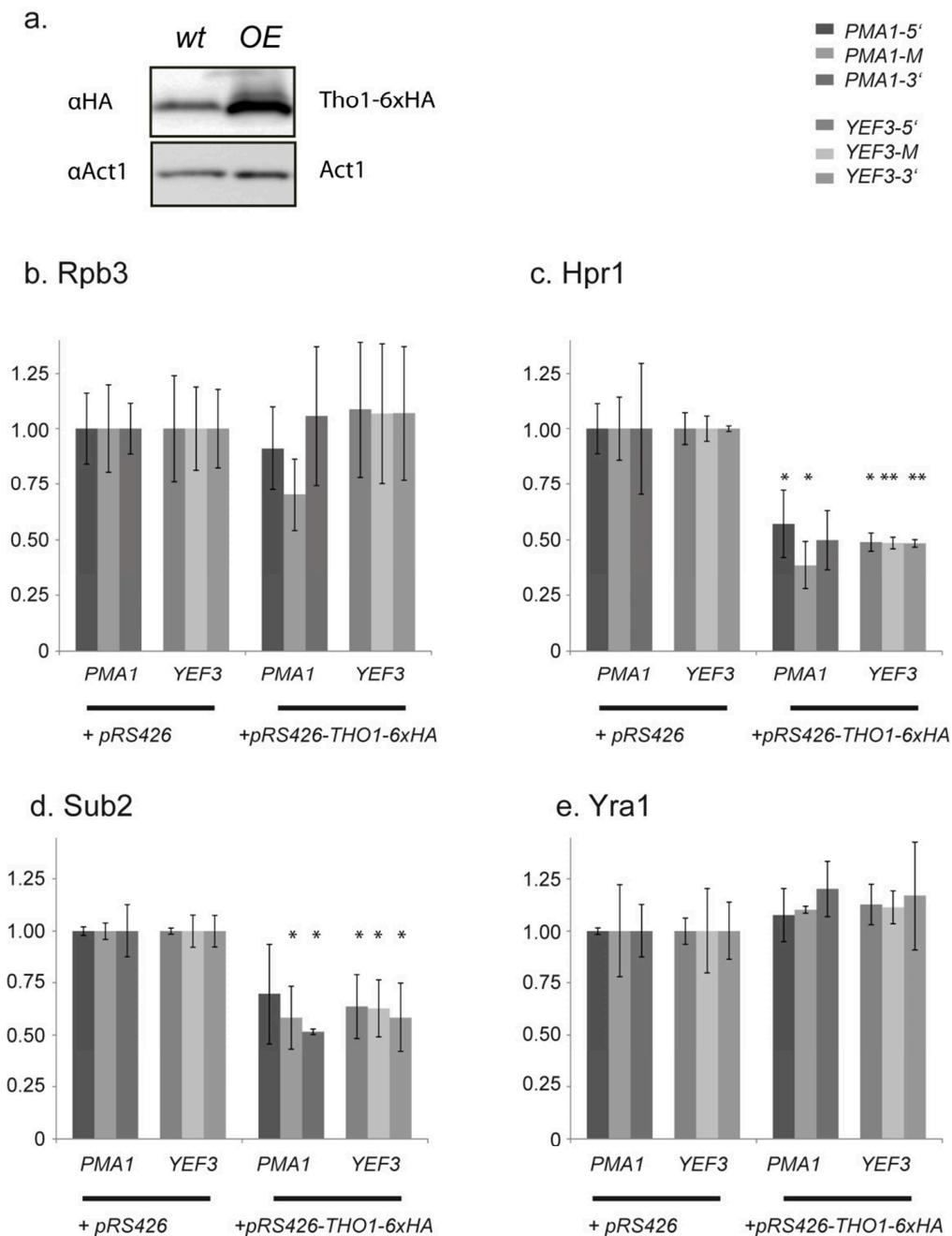


Figure 38: Overexpression of Tho1 leads to decreased THO and Sub2 levels.

(a) Tho1-6xHA protein levels were assessed by Western blotting. The wt strain carried a genomically C-terminal 6xHA tagged *THO1* gene under the control of the endogenous promoter and was assumed to reflect the levels of the untagged protein. The wt strain carried the empty *pRS426* plasmid and the overexpression (OE) strain carried the high copy *pRS426-THO1-6xHA* plasmid additionally to the genomically 6xHA-tagged *THO1*. (b) RNAPII occupancy at *PMA1* and *YEF3* was determined using ChIP, in the wt strain carrying the empty *pRS426* plasmid and in the *THO1*-overexpression strain (*pRS426-THO1-6xHA*). ChIP levels were normalized by the wildtype level of each gene. (c) THO levels (Hpr1) for *PMA1* and *YEF3* in the overexpression strain were decreased. (d) Sub2 levels were decreased in the overexpression strain, while (e) Yra1 levels remained on wt level on *PMA1* and *YEF3*. Results of at least 3 independent experiments are shown (mean +/- SD; *: $p < 0.05$, **: $p < 0.01$, significance relative to the according enrichment in the wt strain).

The results obtained for Tho1 indicate that Tho1 might form together with THO/TREX a functional entity in the cell, since the occupancies of Tho1 and THO/TREX correlate very well with each other. This cannot be solely explained by recruitment to active chromatin, since Tho1 correlation is much worse to RNAPII (Rpb3) as to TREX. Moreover, experiments with Spt5 mutants indicate a role for the Spt5-CTR in recruiting Tho1 to the transcription site. Interestingly, THO/TREX behaves different to Tho1 upon Spt5-CTR mutation, possibly indicating a regulation of THO/TREX by Tho1. Overexpression of Tho1 resulted in a decrease of THO and Sub2 levels at active chromatin, while Yra1 levels stayed at *wt* levels. This shows that Tho1 can modulate TREX levels at actively transcribed genes, but the mechanism and the biological function remain to be shown.

III. Discussion

The TREX complex is essential for gene expression in the cell. Although over the last years several functions for the TREX complex have been shown, it remained enigmatic how and where on a genome-wide level the TREX complex is recruited.

1. The mRNA export receptor adapters TREX, Nab2 and Npl3 are recruited to all genes

By genome wide high resolution ChIP-chip occupancy mapping, we showed that the multifunctional TREX complex is recruited to all protein coding genes. Interestingly, we found that the recruitment is length dependent, *i.e.* increases from 5' to 3' throughout the gene in an RNA independent manner (Figure 6, Figure 13, Figure 14). This recruitment mode is by now unique to the THO/TREX complex, as we could not detect similar characteristics for general elongation factors or for other mRNP biogenesis factors like *e.g.* Nab2 and Npl3 (Figure 7 and Figure 16). Furthermore, we found that not all TREX subunits behave exactly the same: Hrb1 and Gbp2, two SR like proteins show very similar behavior at the 5' of the gene, but reach after *app.* 400 bp a plateau, which slightly declines to the 3' end of the gene. The levels of Gbp2 and Hrb1 vanish at *app.* the polyA site, exactly as the other TREX subunit levels do (Figure 7). This indicates that TREX is recruited as one complex, but that not the whole complex stays associated, but that Hrb1 and Gbp2 appear to be more dynamic in their binding and are supposedly transferred to the nascent RNA. Also for Yra1, we observed variations from strictly THO similar recruitment: Yra1 was recruited to much lower amounts to intron containing genes, as expected by THO recruitment, however, Yra1 was still present at intron containing genes. Whether this is due to the recruitment of a different complex or whether this is due to dissociation of Yra1 from TREX after recruitment of the whole complex can only be speculated, however, the latter seems more likely, since, *e.g.* a transfer of Yra1 away from TREX could fully explain this phenomenon and would be in line with the model proposed by Abruzzi et al. (2004).

The occupancy profiles of the “non-TREX” mRNA export receptor adapters Nab2 and Npl3 look very much like the profiles of Gbp2 and Hrb1 (Figure 7). They are also recruited during the elongation phase as expected for mRNA export factors and also leave the site of transcription at *app.* the poly adenylation and cleavage site. This seems to be a characteristic for all assayed mRNA export receptor adapters. Nab2 and Npl3 exhibit a more plateau like shape which declines slightly from 5' to 3'. This indicates a different recruitment mode for Nab2 and Npl3 compared to TREX. Dermody et al. (2008) showed that Npl3 is recruited to the genes by binding to the Rpb1-CTD and then -upon

phosphorylation by CK2 (Casein Kinase 2) on Ser411 - Npl3 is handed over to the mRNP. The transfer to the mRNA might explain why Npl3 levels do not constantly increase, although Ser2 phosphorylation levels of the CTD constantly increase from 5' to 3', since the mRNA bound Npl3 cannot be detected by ChIP over a long range (see also Figure 19). However, we observed all TREX subunits, Npl3 and Nab2 at all actively transcribed genes, indicating that at least at the stage of transcription different mRNA export pathways do not exist.

Interestingly, Nab2 was strongly cross-linked to RNAPIII genes, which was unexpected at that moment. High signals, which are expected by high activity of the RNAPIII genes and the strong Pearson correlation to the RNAPIII largest subunit signals in ChIP-chip lead to the assumption that Nab2 might be involved in RNAPIII transcription or early maturation of RNAPIII transcripts. To further characterize the role of Nab2 in RNAPIII metabolism, a follow up project was started by Max Reuter in the Sträßer lab, to elucidate the role of Nab2 in RNAPIII transcription. Recently, also Gonzalez-Aguilera et al. (2011) published indications for an involvement of Nab2 in RNAPIII metabolism: overexpression of Nab2 genetically interacts with *CRM1* and *UTP8*, which are both involved in RNAPI and III metabolism. Furthermore, they showed that in a Nab2 degon strain tRNA metabolism genes are down regulated among other genes indicating an involvement of Nab2 in tRNA biogenesis. Furthermore, inspection of the CRAC data, published by Tuck and Tollervey (2013), also revealed that Nab2 binds to tRNAs and other RNAPIII transcripts (*RPR1* and *SCR1*) during transcription or very early after transcription, since binding was found to the full length transcripts that are not yet processed by trimming. Therefore two scenarios or even a combination of both seems likely: First, Nab2 could, similar as in the process of RNAPII transcription, promote elongation of RNAPIII *in vivo* (Gonzalez-Aguilera et al. 2011a). Second, Nab2 could carry out a function in a downstream process on the resulting RNA which takes place very early after its co-transcriptional recruitment.

2. Recruitment of the TREX complex

How TREX is recruited to the transcription site and how it interacts with the RNAPII has remained unknown for several years. In this study we show that THO/TREX increases from 5' to 3' in the genome wide recruitment profiles. Furthermore, we show by a novel ribozyme based assay that all TREX subunits and the two mRNA export receptor adaptors Nab2 and Npl3 depend on the nascent mRNA for efficient recruitment *in vivo* (Figure 17 and Figure 18). As a caveat, the efficiency of ribozyme cleavage in this context is not known, prohibiting quantitative conclusions of the influence of RNA to the recruitment. The RNA dependent recruitment of the whole TREX complex is in contrast to an earlier study (Abruzzi et al. 2004), which showed that TREX components Sub2, Yra1 and Hpr1,

have varying degrees of RNA dependent interaction with chromatin. Because our ribozyme cleavage assay affects each of these complex members equivalently, we suggest that discrepancies within this previous study are due to the use of RNase digestion to assess RNA-dependent recruitment. More specifically, because nuclease digestion follows the formaldehyde crosslinking steps, hypothetically this treatment may result in RNase-resistant interactions that were RNA dependent *in vivo*. Hence, the newly established ribozyme assay reflects the situation in the cell more accurate since the cleavage takes place before crosslinking.

Importantly, we used the ribozyme assay to show that the 5' to 3' increase of THO/TREX is not mediated by the nascent mRNA, since the RNA length in the ribozyme assay does not correlate with the TREX signal further downstream of the cleavage site (Figure 19).

Furthermore, we provide strong indication that the by now unique length dependent recruitment mode of THO/TREX is mediated by direct interaction with phosphorylated Rpb1-CTD. Tyr1 and Ser2 phosphorylation marks show a similar 5'-3' increasing profile during elongation and correlate highly with THO/TREX (Figure 6 and Figure 20). CTD mutants revealed that correct phosphorylation of the CTD is necessary for efficient THO/TREX recruitment *in vivo* (Figure 21 and Figure 22). Remarkably, analysis of CTD mutants indicated that Tyr1 and Ser2 phosphorylation are interdependent *in vivo*. As a caveat, this effect could also be an artifact due to changes nearby the epitope of the antibody. However, a similar interdependence is known for the Ser5 and Ser2 phosphorylation. Ser5 phosphorylation, which peaks during initiation phase of transcription, was shown to recruit Bur1 kinase, which in turn carries out initial Ser2 phosphorylation to prepare the CTD for the major Ser2 kinase Ctdk1 (Cho et al. 2001, Qiu et al. 2009). The interdependence between Ser2 and Tyr1 phosphorylation might be due to at least three reasons: Firstly, the Tyr1/Ser2 kinases need the presence of the Ser2/Tyr1 phosphorylation for their recruitment or activity. Secondly, Ser2 phosphorylation leads to inhibition of Tyr1 phosphatase/s and Tyr1 phosphorylation to inhibition of the Ser2 phosphatase/s. Thirdly, the decrease of Ser2 or Tyr1 phosphorylation impairs the whole transcription machinery, which accumulates such severe secondary effects in the CTD mutants that the CTD phosphorylation becomes defective. However, the last reason seems unlikely, since in such a case severe effects of the CTD-mutant on RNAPII occupancy at the genes would be expected, which were not observed (Figure 21). Hence, the possible interconnection of Ser2 and Tyr1 phosphorylation might be helpful for the discovery of the unknown Tyr1 kinase or phosphatase, since these should be connected to the Ser2 phosphorylation of the CTD.

In a next step, we showed direct binding of THO to the Rpb1-CTD *in vitro* by pull-down assays, which indicate weak interaction of pure THO complex with Ser2 phosphorylated and Ser5 phosphorylated CTD repeats. The strongest interaction was detected between THO and Ser2-Ser5 diphosphorylated

peptides. Interestingly, Tyr1 phosphorylated peptides showed low to none binding, possibly pointing out towards a negative regulation of THO binding by Tyr1 phosphorylation. This binding pattern is in line with the published ChIP-chip data for the CTD phosphorylations: The Ser2P antibody, used for the ChIP-chip experiments, recognizes at least *in vitro* only Ser2 phosphorylated and the combination of Ser2-Ser5 diphosphorylated CTD peptides and not Tyr1-Ser2 diphosphorylated peptides (Chapman et al. 2007), indicating that indeed Tyr1-phosphorylation-free Ser2 phosphorylated CTD-repeats for THO/TREX binding should be available and that their phosphorylation pattern is also increasing throughout gene length.

The recruitment of THO/TREX by Ser2-Ser5 diphosphorylated CTD raises a new, important question: How is THO/TREX again released from the transcription site? The presented ChIP-chip data clearly show that TREX leaves the gene at the polyadenylation and cleavage site, suggesting its dissociation from RNAPII before transcription termination (Figure 6 and Figure 7). The decrease of TREX falls together with the release of the mRNP from the transcription site. This dissociation of TREX could be brought about by the decrease in Ser2-Ser5 phosphorylation at *app.* the polyadenylation site. To answer whether the Ser2-Ser5 profiles are sufficient to explain the release of TREX, we tried to simulate Hpr1 occupancy at the genes by fitting with the Ser2 and Ser5 phosphorylation profiles. We were not able to simulate the release by the combination of both occupancies. This could be due to two reasons: Firstly, it is not known which exact combinations of phosphorylation occur and at which stage of the transcription cycle they occur. Hence, most likely double phosphorylations cannot be simply estimated by combination of the two single occupancies for Ser2 and Ser5 phosphorylation, especially since it was shown that both are interdependent, as Ser5 phosphorylation is involved in recruiting Bur1 kinase, which phosphorylates Ser2 (Qiu et al. 2009).

As a second reason, an additional factor/event might be needed to release TREX from the transcription site: TREX could be dissociated by 3'end processing factors that bind to the CTD when Tyr1 phosphorylation decreases and polyadenylation factors are recruited (Mayer et al. 2010, Mayer et al. 2012a, Hsin et al. 2011). This seems very likely since the meta gene occupancy profiles of TREX components more closely resemble that of Tyr1P than Ser2P (Figure 3A) but THO does not bind to the Tyr1 phosphorylated CTD (Figure 27). Furthermore, an involvement of an 3'end processing factor like *e.g.* Pcf11, which was shown to interact with Yra1, or Rna14/15 in the dissociation of THO/TREX seems very supposable, since the occupancy of these 3'end processing factors was shown to peak exactly when THO/TREX occupancy strongly decreases (Figure 40). Furthermore, an involvement of the 3'end processing factors is supported by the fact that extensive coupling of 3'end processing with THO/TREX was reported, *e.g.* $\Delta mft1$ (THO) and *sub2-201* mutant cells (TREX) display faulty 3'end processing with defective release of the mRNP from the chromatin (Rougemaille et al. 2008).

Furthermore, the cleavage of the mRNA itself, could also contribute to TREX dissociation since TREX depends on the nascent RNA for efficient recruitment (Figure 18). Additionally, TREX dissociation could be enhanced by loss of Prp19C function, as Prp19C was shown to be needed to stabilize THO/TREX recruitment at the 3' end of the gene (Chanarat et al. 2012). Hence, most likely multiple processes ensure the timely dissociation of the proteins from the transcription site and thereby regulate one of the hallmarks in early mRNP formation, the release of the mRNP from the chromatin. Sub2 and Yra1 might thereby leave the transcription machinery as part of the mRNP, whereas THO might either bind to the mRNP or be directly recycled for a new round of transcription. The whole model is summarized in Figure 39.

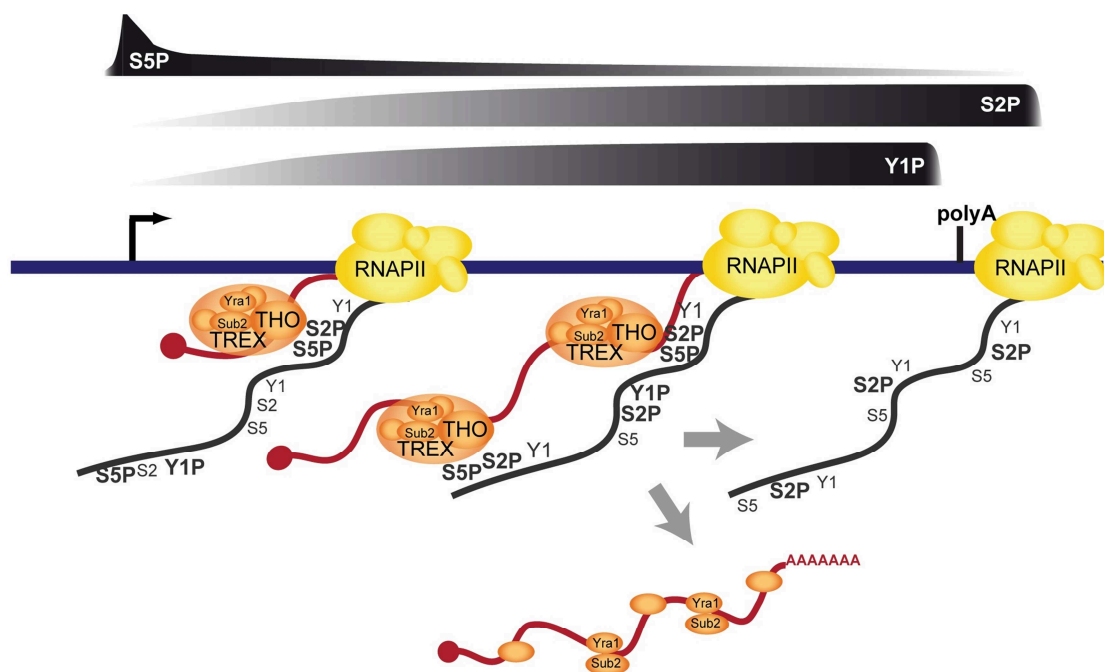


Figure 39: Model of TREX recruitment and mRNP formation

In the upmost panel, Ser5, Ser2 and Tyr1 phosphorylation are represented, as the levels were determined by ChIP-chip. Below, an average yeast gene is depicted in blue with promoter (black arrow) and cleavage and polyA-site (polyA). The yellow RNAPII is depicted in several states of the transcription cycle with a CTD carrying three exemplary repeats with the according phosphorylations (black). The CTD is during initiation (leftmost) Ser5 phosphorylated and as the polymerase traverses through the gene more and more Tyr1 and Ser2 is phosphorylated, while Ser5 decreases to a basal level during elongation phase. As soon as Ser2-Ser5 diphosphorylated CTD peptides occur, TREX is recruited, via direct binding of THO to the CTD. TREX can bind simultaneously to the CTD and the RNA and thereby offers the cell the possibility to keep the growing RNA in spatial vicinity to the CTD, so that CTD-binding mRNP biogenesis factors (e.g. capping enzyme, splicing proteins etc.) can carry out their function efficiently. At the cleavage and polyA site, TREX is dissociated and the release of the mRNP takes place as well as polyadenylation. This event might be regulated by the loss of Tyr1 phosphorylation and/or by thereby recruited 3' end processing factors, like e.g. *Pcf11*.

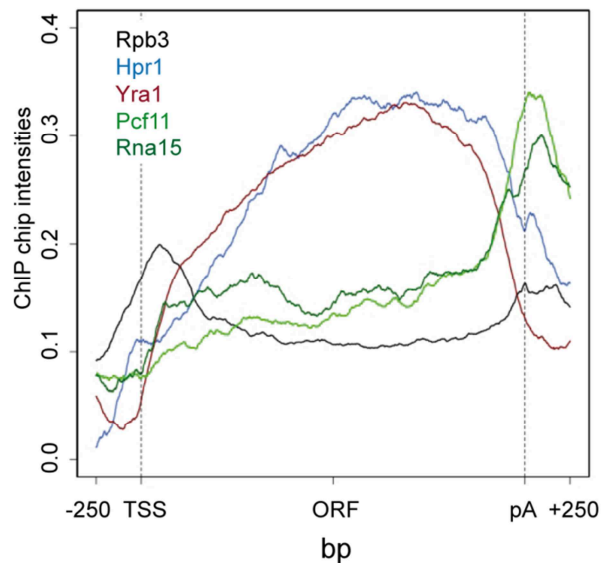


Figure 40: Recruitment profiles of RNAPII, THO/TREX and Pcf11

Recruitment meta-profiles of RNAPII (Rpb3), THO/TREX (Hpr1 and Yra1) as well as the 3' end processing factors Pcf11 and Rna15 (occupancies from Mayer et al. 2010 and Mayer et al. 2012b) are shown. Interestingly, Pcf11 and Rna15 peak exactly where THO/TREX signals strongly decrease. Pcf11 binding was shown to be antagonized by Tyr1 phosphorylation of the Rpb1-CTD (Mayer et al. 2012a).

3. Is the recruitment of TREX to the phospho-CTD of Rpb1 conserved?

Even though it is highly speculative to assume a similar recruitment mechanism in higher eukaryotes, the interaction of THO with the phospho-CTD might be conserved in metazoans. Such a direct interaction of TREX with the transcription machinery might be the basis for the TREX dependent export of naturally intronless mRNAs, like *e.g.* HSPB3, IFN- α 1 and IFN- β 1 (Lei et al. 2011). Recently, a sequence element termed CAR-E that promotes association of TREX with intronless mRNAs was described (Lei et al. 2013). A direct interaction of TREX with the transcription machinery and then transfer of TREX to the RNA, similar as it has been reported for many proteins in mRNP biogenesis, *e.g.* for Npl3, Rna15, *etc.*, might be the basis for TREX recruitment to intronless RNAs.

However, after recruitment of TREX to transcription site of intron containing genes, TREX could help to recruit the splicing machinery, consistent with the largely cotranscriptional splicing. Noteworthy in this context, the human transcription factor CA150, which binds to the phosphorylated CTD, interacts directly with the splicing factor SF1 repressing transcription elongation, giving an example for a transcription factor recruiting a splicing factor indirectly to the CTD (Carty et al. 2000, Goldstrohm et al. 2001). However, if direct interaction of THO/TREX with the Rpb1-CTD is not conserved, metazoan TREX could be recruited to the transcription machinery by the spliceosome, shifting the recruitment mode from direct CTD interaction to indirect CTD interaction. Such an indirect CTD interaction could be mediated by one or more subunits of the spliceosome or splicing factors. A possible candidate for

mediating such an indirect interaction would be U2AF65, which was shown to directly interact with hyperphosphorylated CTD (David et al. 2011). Furthermore, U2AF65 was shown to recruit the Prp19C to the transcription site (David et al. 2011). Recently, it was shown that at least in yeast the Prp19C ensures full occupancy by binding to TREX in the 3' end of the genes (Chanarat et al. 2011). However, it will be interesting to see whether TREX recruitment to the nascent mRNA also increases towards the 3'-end of the gene in higher eukaryotes and, if so, whether TREX will be important for the expression of long transcripts in higher eukaryotes similar as in yeast.

4. Possible implications of increasing TREX occupancy for mRNP biogenesis

The 5' to 3' increase in occupancy of core THO/TREX throughout the gene seems to be a unique feature for a transcription factor. Exploiting a hypomorphic mutant of *THO2* we show that the 5'-3' increase in TREX occupancy is important for the correct expression of long genes. THO, Sub2 and Yra1 might be needed at higher levels towards the 3' end of genes, to keep the nascent mRNA in the vicinity of the CTD (as described above and shown in Figure 39). Hypothetically, this could be important to ensure efficient and correct processing and packaging of the mRNA, which is consistent with the finding that a continuous transcript is needed for mRNA processing (Fong et al. 2009). In this model, TREX would ensure by bridging between the nascent mRNA and the Rpb1-CTD spatial proximity of CTD recruited mRNP biogenesis factors to the RNA, their target of action. All hallmarks in the early mRNA maturation and mRNP formation are known to be regulated by the CTD, *e.g.* the capping enzyme subunit Ceg1 is recruited to Ser5 phosphorylated CTD (Cho et al. 1997) or Prp40, a subunit of the U1 snRNP, which is involved in splicing, was shown to be recruited to Ser2 phosphorylated CTD (Morris et al. 2000). It has been shown that splicing reactions *in vitro* are more efficient in presence of recombinant and hyperphosphorylated CTD (Millhouse and Manley 2005). Furthermore, several mRNA binding proteins, like *e.g.* Npl3 are also recruited to the CTD (Dermody et al. 2008) and subsequently handed over to the mRNP. Moreover, the 3'end processing machinery, which regulates cleavage and polyadenylation, is also recruited to the Ser2 phosphorylated CTD (Noble et al. 2005, Barilla et al. 2001, Kyburz et al. 2003, Dichtl et al. 2002). Thus a bridging protein complex would give the cell the opportunity to enhance efficiency of such CTD-coupled mRNP maturation processes, by ensuring a sufficient window of opportunity for each step.

If the RNA over its whole length is tethered to the Rpb1-CTD or only parts of the nascent RNA are tethered to the CTD by TREX remains thereby as an open question. However, even short periods of time, in which the RNA is in close proximity to the CTD, could extend the window of opportunity and thus enhance efficiency tremendously and prevent defective mRNP formation. A further question

arising from this model would be how the RNA tethering could be organized on a molecular basis: In theory, a fully extended CTD is approximately 700 Å long in yeast (Meinhart et al. 2005), which corresponds to the length of a 2.5 kb long linear mRNA. Since the median length of an mRNA in *Saccharomyces cerevisiae* is 1436 nucleotides (Nagalakshmi et al. 2008b), the CTD is in principle able to span the entire length of an average mRNA. However, it seems unlikely that the CTD as well as the mRNA exist in a fully extended form *in vivo*. Thus, it remains to be elucidated, how mRNP formation is spatially organized. The RNA itself could *e.g.* form several loops in a way that several points of the RNA are associated with the CTD. This model would also be highly compatible with the fact that mRNAs in higher eukaryotes usually carry several large introns, which are cleaved out in a lariat like loop structure and by the fact that the human CTD is twice as long as the yeast CTD. In such a model, TREX could even promote splicing by helping to establish the lariat like loops for the splicing reaction. Hence, we propose that TREX promotes mRNP packaging through its bifunctional binding to the CTD and RNA by ensuring spatial proximity of the nascent mRNA to mRNP binding proteins, which are recruited to the CTD. In summary, we identify a direct interaction of TREX with the phospho-CTD as a molecular mechanism of TREX recruitment to transcribed genes (Figure 39). Thus, in addition to its many known functions the CTD code probably also coordinates transcription with mRNA export.

5. Gbp2 and Hrb1 may be recruited by THO to the CTD and then transferred to the mRNA

Interestingly, we observed in the distribution shapes over the genes differences between the two SR-like proteins Hrb1/Gbp2 and the other THO/TREX subunits. The profiles of Gbp2 and Hrb1 recapitulate the one of Npl3, a third yeast SR-like protein. Additionally, their peak occupancies correlate well, *i.e.* Pearson correlation coefficients for Gbp2 and Hrb1 to Npl3 are higher as compared to THO/TREX (Figure 7 and Figure 15). Npl3 was proposed to be recruited to the transcription site by the CTD and then transferred to the mRNA (Dermody et al. 2008). We hypothesize that Gbp2 and Hrb1 are also transferred from the transcription site, to which they are recruited via THO/TREX, to the mRNA. Noteworthy in this context, Hrb1 and Gbp2 are in their domain architecture very similar to Npl3 (depicted in Figure 41). All three proteins comprise the SR-like protein typical two RRM domains, which usually interact directly with mRNA. The C-termini of Hrb1 and Gbp2, which are less conserved to the C-terminus of Npl3, but share high similarity among each other most likely do not bind like the C-terminus of Npl3 to the Rpb1-CTD. They could both serve for binding of Gbp2/Hrb1 to TREX. Furthermore, the N-termini of Gbp2 and Hrb1 are also very similar and both contain a RS rich region. Such regions are known from SR proteins to often mediate protein-protein interactions (Graveley and Hertel) and could therefore also serve as an interface to

TREX. This would lead to a recruitment of Hrb1 and Gbp2 to the transcription site via TREX and thereby indirectly by the CTD. Subsequently, Hrb1 and Gbp2 could be –analog to Npl3 after its recruitment by the CTD- handed over to the mRNP, to which they supposedly bind with their conserved middle two RRM domains. As soon as Gbp2 and Hrb1 would be handed over to the RNA, no length dependency could be detected anymore, since the growing mRNA chain takes the proteins further and further away from the transcription site and their crosslink-ability to the DNA would be thereby reduced. An interesting question is, if Gbp2 and Hrb1 also need a posttranscriptional modification, similar to the phosphorylation of Npl3 on S411 (Dermody et al. 2008), which lies within the CTD interaction domain, to regulate their transfer from the transcription site to the mRNA.

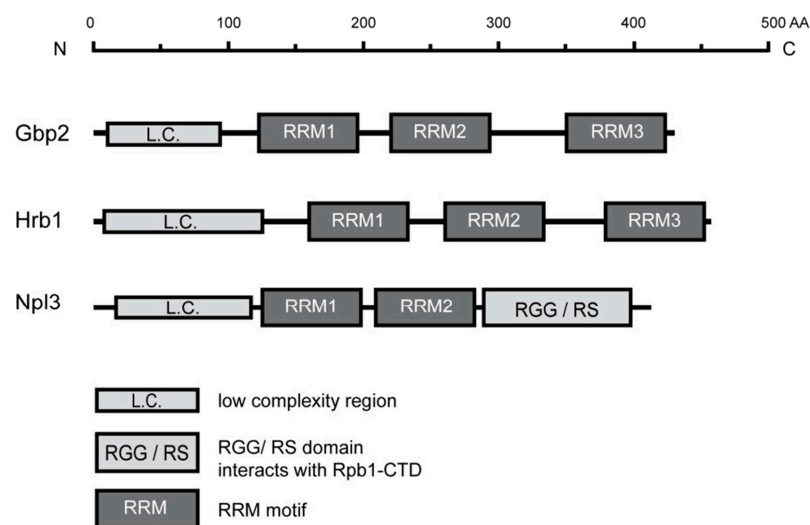


Figure 41: predicted domain architecture of Gbp2, Hrb1 and Npl3

The domains are drawn to scale as annotated by SMART. The RGG/RS domain which is known to interact with the Rpb1-CTD is mapped by SMART as a low complexity region, but identified by (Dermody et al. 2008) as Rpb1 interaction domain with RGG / RS repeats. RRM3 in Gbp2 and Hrb1 share very high sequence similarity. Also the both low complexity regions of Gbp2 and Hrb1 show very high sequence similarity and are RS rich.

6. The Spt5-CTR: a possible role in mRNP biogenesis and export

Similar to the CTD of the largest subunit of RNAPII, Spt5 also has a repetitive and most likely unstructured C-terminal region (CTR). Spt5 is a general elongation factor and is thought to be attached very tightly to the core of RNAPII. Furthermore, Spt5 was found by ChIP-chip to be generally present at all RNAPII genes (Mayer et al. 2010) and could therefore serve as a general recruitment platform in transcription. Recently, the Paf1 complex (Paf1C) was shown to depend on correct phosphorylation of the Spt5-CTR for efficient Paf1C recruitment to the transcription site (Qiu et al. 2012b). Moreover, (Mayer et al. 2012b) showed that Spt5-CTR is needed for efficient 3' end processing and also that the CTR contributes to stable recruitment of the capping enzyme

(Lidschreiber et al. 2013). This already points out a presumable role of the Spt5-CTR in promoting mRNP biogenesis. Interestingly, (Lindstrom et al. 2003) identified by co-purification and mass-spectrometry analysis, that several proteins involved in mRNP biogenesis copurify with Spt5-FLAG, among those were capping enzyme subunits, 3' end processing factors and also Tho1. By employing Spt5-CTR mutants abolishing phosphorylation (Figure 36), we could show that Tho1 recruitment to the transcription site is dependent on the CTR, while a phospho-mimicry mutation recovers Tho1 levels almost completely, indicating at least indirect involvement of the Spt5-CTR in Tho1 recruitment (Figure 37). Curiously, we found that the Ser1 to Ala mutants, which decreased Tho1 recruitment, enhanced THO/TREX occupancy, especially at the 3' end, while the phospho-mimicry mutant lowers the levels of THO/TREX, and almost completely restored Tho1 levels. Thus, we describe for the first time a difference in the recruitment between Tho1 and THO/TREX. This is of special interest, since the genome wide recruitment profiles are very similar and Tho1 depends on THO/TREX in its recruitment (Figure 35), indicating that both are usually recruited together and/or form a functional entity. Both possibilities are consistent with the data from higher eukaryotes, which indicate, that CIP29, the homologues protein of Tho1 in human, copurifies ATP-dependently with TREX (Dufu et al. 2010). The effect of the Spt5-CTR mutants could be explained by a regulatory role of Tho1 on THO/TREX during transcription, if Tho1 binding to Spt5 induces a regulatory function of Tho1 on THO/TREX. In line with this hypothesis, overexpression of Tho1 leads to a diminished occupancy of THO (Hpr1) and Sub2 on the genes but not of Yra1. A possible explanation for this effect might be that Tho1, similar to its homologous protein CIP29, is able to stimulate its' DEAD box helicase activity. In particular, CIP29 was shown to stimulate DDX39, one of the two homologues of Sub2 in higher eukaryotes (Sugiura et al. 2007). Whether Tho1 also stimulates Sub2 helicase activity in yeast and thereby helps to remodel the early mRNP and also dissociates THO and Sub2 from the transcription site and early mRNP, remains to be shown and is ongoing work in collaboration with the group of Elisabeth Tran (Purdue University, Indiana, US).

Overall, a role of Spt5-CTR in mRNP biogenesis and most likely in regulating early steps of mRNA export via Tho1 and TREX seems very likely, but details remain to be elucidated in the future.

7. Conclusions

In this study we provide novel insights into the genome-wide recruitment of mRNA export factors, and show that all mRNA export receptor adaptors are recruited to all RNAPII genes. Additionally, our data suggest a new function for the polyA binding protein Nab2 in RNAPIII metabolism and it will be interesting to see what exactly the function of Nab2 is in this context. Furthermore, we could

elucidate the molecular mechanism for the recruitment of TREX: all TREX subunits depend in their recruitment on the nascent mRNA. Moreover, direct binding of the THO subcomplex of TREX to the phosphorylated Rpb1-CTD explains most likely the new and unique recruitment mode we observed for TREX. The 5' to 3' increasing occupancy of TREX is physiological relevant for the cell and might imply a novel mechanism in mRNP formation: The bifunctional TREX complex could bridge between the CTD and the nascent mRNA to promote effective mRNP biogenesis. It will be fascinating to determine the impact of this special recruitment mode and the possible bridging of TREX to the CTD on mRNP formation. Additionally, to unravel the mechanism releasing TREX from the transcription site, will also be an important question in the future. A possible candidate for a releasing mechanism, apart from the cleavage of the mRNA and the 3'end processing factors, is Tho1, a protein which associates with TREX, but upon overexpression down regulates TREX occupancy at the genes. Furthermore, it will be exciting to see, if the recruitment of THO/TREX by the phosphorylated CTD of Rpb1 is conserved in higher eukaryotes, since molecular details of TREX recruitment in presence or absence of introns are still not known in metazoans.

IV. Material and Methods

1. Materials

1.1 Consumables and Chemicals

Consumables and chemicals were purchased from the following companies: Affymetrix (SantaClara, CA), Applichem (Darmstadt), Applied Biosciences (Darmstadt), Beckman Coulter (Krefeld), Biomol (Hamburg), Bio-Rad (Munich), Biozym (Hess. Oldendorf), Chemicon (Temecula, Canada), Diagenode (Liege, Belgium), Eppendorf (Hamburg), Fermentas (St. Leon-Rot), Formedium (Norwich, UK), GE Healthcare (Munich), Gilson (Bad Camberg), Invitrogen (Karlsruhe), Life Technology (Darmstadt), Macherey&Nagel (Düren), Medac (Hamburg), MembraPure (Bodenheim), Merck Biosciences (Darmstadt), Millipore (Molsheim, France), Mobitec (Göttingen), NEB (Frankfurt), Neolab (Heidelberg), Nunc (Wiesbaden), PANAtecs (Tübingen), Peske (Aindling-Arnhofen), Promega (Mannheim), PSL (Heidelberg), Qiagen (Hilden), Roche (Mannheim), Roth (Karlsruhe), Sarstedt (Nümbrecht), Semadeni (Düsseldorf), Serva (Heidelberg), Sigma (Taufkirchen), Stratagene (Amsterdam, Netherlands), Thermo Scientific (Munich), VWR (Ismaning)

1.2 Equipment

Table 1: Equipment used in this study

Name	Supplier
GeneChip Fluidics Station 450	Affymetrix (SantaClara, CA)
GeneChip Scanner 3000 7G	Affymetrix (SantaClara, CA)
Bioruptor UCL 200	Diagenode (Liege, Belgium)
Beckman DU650 spectrophotometer	Beckman Coulter (Krefeld)
L80 and Optima™ L-90 K ultracentrifuge	Beckman Coulter (Krefeld)
SW32, SW40 rotor	Beckman Coulter (Krefeld)
CO8000 Cell Density Meter	WPA (Cambridge, UK)
Dissection microscope manual MSM	Singer (Somerset, UK)
Electrophoresis Power Supply Consort E853	Neolab (Heidelberg)
Heidolph shaker duomax 1030	Neolab (Heidelberg)
Rotator, Vortex Genie2	Neolab (Heidelberg)
Mini-Protean II system	Biorad (Munich)
ND-1000 Spectrophotometer	Thermo Fisher Scientific (Munich)
Innova 44 shaking incubator	New Brunswick Scientific (Nürtingen)
Optimax TR developing machine MS	Laborgeräte (Dielheim)
Pulverisette	Fritsche (Idar-Oberstein)
Research Pipettes P2, P20, P200 P1000	Gilson (Bad Camberg)
Rotanda 46R, 460R	Hettich (Tuttlingen)

Sorvall Evolution RC, RC 5B Plus	Thermo Fisher Scientific (Munich)
SLC 6000, GS3, SW34 rotor	Thermo Fisher Scientific (Munich)
StepOnePlus™ Real Time PCR System	Applied Biosystems (Darmstadt)
T3 Thermocycler	Biometra (Göttingen)
Thermomixer compact, BioPhotometer	Eppendorf (Hamburg)
Eppendorf centrifuge 545D, 541R	Eppendorf (Hamburg)
Universal Analytical Balance	Satorius (Göttingen)

1.3 Yeast strains

Table 2: Yeast strains used in this study

Yeast strain	Genotype	Reference
BY4741	MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	Euroscarf
RS453	MAT a; <i>ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i>	(Strässer and Hurt 2000)
<i>RPB3-TAP</i>	<i>RPB3-TAP::HIS3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>HPR1-TAP</i>	<i>HPR1-TAP::URA3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>TAP-THO2</i>	<i>TAP-THO2, MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>THO2-TAP</i>	<i>THO2-TAP::URA3, MATa; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i>	this study
<i>MFT1-TAP</i>	<i>MFT1-TAP::URA3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>GBP2-TAP</i>	<i>GBP2-TAP::URA3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>HRB1-TAP</i>	<i>HRB1-TAP::URA3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>TAP-NPL3</i>	<i>TAP::NPL3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>NAB2-TAP</i>	<i>NAB2-TAP::URA3; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>HPR1-TAP MFT1-His₇</i>	<i>HPR1-TAP::TRP1; MFT1-myc-His₇::HIS3; MAT alpha; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i>	this study
<i>RIX1-TAP</i>	<i>RIX1-TAP::TRP1; MAT alpha; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i>	this study
<i>PCF11-TAP</i>	<i>PCF11-TAP::TRP1; MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i>	this study
wt, σ active	<i>Ribo-YCT1::KanMX ; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
wt, σ inactive	<i>RiboMut-YCT1::KanMX ; MATa; his3Δ1; leu2Δ0; met15Δ 0; ura3Δ0</i>	this study
<i>RPB3-TAP</i> σ active	<i>RPB3-TAP::HIS3; Ribo-YCT1::KanMX; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>RPB3-TAP</i> σ inactive	<i>RPB3-TAP::HIS3; RiboMut-YCT1::KanMX ; MATa; his3Δ 1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>HPR1-TAP</i> σ active	<i>HPR1-TAP::URA3, Ribo-YCT1::KanMX; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>HPR1-TAP</i> σ inactive	<i>HPR1-TAP::URA3, RiboMut-YCT1::KanMX; MATa; his3Δ 1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>TAP-THO2</i> σ active	<i>TAP-THO2, Ribo-YCT1::KanMX ; MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>TAP-THO2</i>	<i>TAP-THO2; RiboMut-YCT1::KanMX; MATa; his3Δ1; leu2Δ 0;</i>	this study

σ inactive	<i>met15Δ0; ura3Δ0</i>	
THP2-TAP σ active	THP2-TAP::URA3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
THP2-TAP σ inactive	THP2-TAP::URA3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
MFT1-TAP σ active	MFT1-TAP::URA3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
MFT1-TAP σ inactive	MFT1-TAP::URA3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
HRB1-TAP σ active	HRB1-TAP::URA3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
HRB1-TAP σ inactive	HRB1-TAP::URA3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
GBP2-TAP σ active	GBP2-TAP::URA3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
GBP2-TAP σ inactive	GBP2-TAP::URA3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
NAB2-TAP σ active	NAB2-TAP::URA3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
NAB2-TAP σ inactive	NAB2-TAP::URA3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
TAP-NPL3 σ active	TAP::NPL3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
TAP-NPL3 σ inactive	TAP::NPL3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
THO1-TAP σ active	THO1-TA::URA3; Ribo-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
THO1-TAP σ inactive	THO1-TAP::URA3; RiboMut-YCT1::KanMX; MATa; <i>his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
YRA1 shuffle	<i>yra1::HIS3; MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+; pRS316-YRA1</i>	(Strässer and Hurt 2000)
YRA1 shuffle HPR1-TAP	<i>yra1::HIS3; HPR1-TAP::TRP1; MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+; pRS316-YRA1</i>	this study
YRA1 shuffle Tho2-TAP	<i>yra1::HIS3; THO2-TAP::TRP1; MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+; pRS316-YRA1</i>	this study
YRA1 shuffle Mft1-TAP	<i>yra1::HIS3; Mft1-TAP::TRP1; MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+; pRS316-YRA1</i>	this study
RPB1 shuffle	<i>rpb1::HIS3; GAL1::YLR454::TRP1; MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+, pRS316-RPB1</i>	this study
RPB1 shuffle HPR1-TAP	<i>rpb1::HIS3; HPR1-TAP::KanMX, GAL1::YLR454::TRP1, MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+, pRS316-RPB1</i>	this study
Hpr1-Avi, pRS315-BirA-NLS	<i>HPR1-Avi::KanMX, MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+, pRS315-BirA-NLS</i>	this study
Hpr1-Avi, Tho2-TAP pRS315-BirA-NLS	<i>HPR1-Avi::KanMX, THO2-TAP::TRP1KL, MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+, pRS315-BirA-NLS</i>	this study
SUB2-Avi, pRS315-BirA-NLS	<i>SUB2-Avi::KanMX, MAT a; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+, pRS315-BirA-NLS</i>	this study
SUB2-Avi,	<i>SUB2-Avi::KanMX, THO2-TAP::TRP1KL, MAT a; ade2-1; his3-</i>	this study

<i>Tho2-TAP</i> <i>pRS315-BirA-NLS</i>	<i>11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i> , <i>pRS315-BirA-NLS</i>	
<i>YRA1-Avi</i> , <i>pRS315-BirA-NLS</i>	<i>YRA1-Avi::KanMX, MAT α; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i> , <i>pRS315-BirA-NLS</i>	this study
<i>YRA1-Avi</i> , <i>Tho2-TAP</i> <i>pRS315-BirA-NLS</i>	<i>YRA1-Avi::KanMX, THO2-TAP::TRP1KL, MAT α; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i> , <i>pRS315-BirA-NLS</i>	this study
<i>RPB1-Avi</i> , <i>pRS315-BirA-NLS</i>	<i>RPB1-Avi::KanMX, MAT α; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i> , <i>pRS315-BirA-NLS</i>	this study
<i>RPB1-Avi</i> , <i>THO2-TAP</i> <i>pRS315-BirA-NLS</i>	<i>RPB1-Avi::KanMX, THO2-TAP::TRP1KL, MAT α; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i> , <i>pRS315-BirA-NLS</i>	this study
<i>Spt5 shuffle</i>	<i>spt5Δ::HIS3, MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	(Qiu et al. 2012a)
<i>Spt5 shuffle</i> <i>THO1-TAP</i>	<i>THO1-TAP::KanMX, spt5Δ::HIS3, MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>Spt5 shuffle</i> <i>Hpr1-TAP</i>	<i>HPR1-TAP::KanMX, spt5Δ::HIS3, MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>Spt5 shuffle</i> <i>Paf1-TAP</i>	<i>PAF1-TAP::KanMX, spt5Δ::HIS3, MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>THO1-6xHA</i>	<i>THO1-6xHA::KanMX; MAT α; ade2-1; his3-11,15; ura3-52; leu2-3,112; trp1-1; can1-100; GAL+</i>	this study
<i>Δhpr1,RPB3-TAP</i>	<i>RPB3-TAP::HIS3; Δhpr1::KanMX; MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study
<i>Δhpr1,THO1-TAP</i>	<i>THO1-TAP::HIS3; Δhpr1::KanMX; MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</i>	this study

1.4 Plasmids

Table 3: Plasmids used in this study

Plasmid	Description	Reference
pY1WT	<i>rpb1-14xwt repeats</i>	(West and Corden 1995)
pY1WT(7)A9(6)	<i>rpb1-9xwt+5xS2A</i>	(West and Corden 1995)
pY1WT(5)F1(9)	<i>rpb1-5xwt+9xY1F</i>	this study
pRS315-3HA-YRA1	3HA-Yra1(1-226)	this study
pRS315-3HA-yra1-ΔPCID	3HA-Yra1-ΔPCID(aa 77-226)	this study
pRibo-Active	<i>GAL1::GFP</i> -active σ ribozyme	this study
pRibo-Inactive	<i>GAL1::GFP</i> -inactive σ ribozyme	this study
pRS315-BirA-NLS	BirA biotin ligase fused to a nuclear localization sequence, His3	(van Werven and Timmers 2006)
pUG-Myc-C-Avitag	pUG carrying tagging cassette with Myc-Tag and Avitag, kanMX flanked by loxP sites, Amp ^R	(van Werven and Timmers 2006)
pRS313, pRS314, pRS315, pRS316, pRS426	<i>E. coli</i> , yeast shuttle vectors	(Sikorski and Hieter 1989)
pYM-N22, pYM-N25, pYM46	PCR tagging toolbox	(Janke et al. 2004)
pSPT5-wt	SPT5-3HA in single copy LEU2 plasmid YCplac111	(Qiu et al. 2012a)
pSPT5-S1A	spt5-S1-15A-HA3 in single copy LEU2	(Qiu et al. 2012a)

	plasmid YCplac111	
pSPT5-S1D	spt5-S1-15D-HA3 in single copy LEU2 plasmid YCplac111	(Qiu et al. 2012a)
pRS426- <i>THO1</i> -6xHA	pRS426 with genomic <i>THO1</i> app. 500 bp upstream and downstream, a sequence was added to the 3' end of <i>THO1</i> just before the stop codon, encoding an 6x HA-tag	this study

1.5 Oligo-nucleotide sequences

Table 4: Oligo-nucleotides used for cloning

Project	Name	DNA-sequence
Ribozyme plasmids	pYM-N25-Sal-fwd	CGTACGCTGCAGGTCGAC
	pYM-N25-Apal-rev	TAGCTAGGGCCCGTTGTTTATGTTTCGGATGTGA
	pYM-LEU-Apal-fwd	TAGCTAGGGCCCatgtctgcccctatgtctgc
	pYM-LEU-XhoI-rev	TATACTCGAGttaagcaaggattttcttaacttcttc
	Sigma-WT-Ribo	TAGAATTCATAAAGGGCGGCATGGTCCCAGCCTCCTCGCTGGCG CCGCCTGGGCAACATGCTTCGGCATGGCGAATGGGACCAAAGA TATC
	Sigma-MUT-Ribo	TAGAATTCATAAAGGGCGGCATGGTCCCAGCCTCCTCGCTGGCG CCGCCTGGGCAACATGCTTCGGCATGGTGAATGGGACCAAAGA TATC
	SigRibWT/Mut-fwd	TATAGAATTCATAAAGGGCGGCATGG
	SigRibWT-rev	TATAGATATCTTTGGTCCCATTCG
	SigRibMut-rev	TATAGATATCTTTGGTCCCATTCACC
	Seq_pYM-N25_fwd	GCGGCCGCCAGCTG
	Seq_pYM-N25_rev	ACCTAAGAGTCACTTTAAATTTG
	Seq-pYM-N25-Ribo	GCTGCTGGTATTACCCA
CTD mutants	CTD_5wt_2F1_fwd	CCGAGCTATAGTCCAACCTTCACCGAGCTATAGTCCAACCTTCACCG AGCTATAGTCCAACCTTCACCGAGCTATAGTCCAACCTTCACCGAGC TATAGTCCAACCTTCACCGAGCTTTTCACCAACATCACCGAGCTTTT CACCAACATCA
	CTD_5wt_2F1_rev	TCGGTGATGTTGGTGAAAAGCTCGGTGATGTTGGTGAAAAGCTC GGTGAAAGTTGGACTATAGCTCGGTGAAGTTGGACTATAGCTCGG TGAAGTTGGACTATAGCTCGGTGAAGTTGGACTATAGCTCGGTG AAGTTGGACTATAGC
	CTD_3wt_4F1_fwd	CCGAGCTATAGTCCAACCTTCACCGAGCTATAGTCCAACCTTCACCG AGCTATAGTCCAACCTTCACCGAGCTTTTCACCAACATCACCGAGC TTTTACCAACATCACCGAGCTTTTCACCAACATCACCGAGCTTTT CACCAACATCA
	CTD_3wt_4F1_rev	TCGGTGATGTTGGTGAAAAGCTCGGTGATGTTGGTGAAAAGCTC GGTGATGTTGGTGAAAAGCTCGGTGATGTTGGTGAAAAGCTCG GTGAAGTTGGACTATAGCTCGGTGAAGTTGGACTATAGCTCGGT GAAGTTGGACTATAGC
	CTD_7xWT_fwd	CCGAGCTATAGTCCAACCTTCACCGAGCTATAGTCCAACCTTCACCG AGCTATAGTCCAACCTTCACCGAGCTATAGTCCAACCTTCACCGAGC TATAGTCCAACCTTCACCGAGCTATAGTCCAACCTTCACCGAGCTAT AGTCCAACCTCA

	CTD_7xWT_rev	TCGGTGAAGTTGGACTATAGCTCGGTGAAGTTGGACTATAGCTC GGTGAAGTTGGACTATAGCTCGGTGAAGTTGGACTATAGCTCGG TGAAGTTGGACTATAGCTCGGTGAAGTTGGACTATAGCTCGGTG AAGTTGGACTATAGC
	seqCTDMuts	GGTGAAGCACCTACATCT
	seqCTDMuts_rev	CATCTGGAATTTTCATTTTC
HA ₃ -Yra1	SacI-Yra1Pro-fwd	aagagctcCCAATTGTTCTGTTTTCTGTT
HA ₃ -Yra1-ΔPCID	XhoI-NcoI-Yra1Pro-rev	cttCTCGAGTTTCcatggTTAGCAGATGTAGGTATTTTC
	3HA-anneal-fwd	cATGTACCCATACGATGTTCCAGATTACGCTTACCCATACGA TGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGC TGGTGCTC
	3HA-anneal-rev	TCGAGAGCACCAGCGTAATCTGGAACATCGTATGGGTAAGC GTAATCTGGAACATCGTATGGGTAAGCGTAATCTGGAACAT CGTATGGGTA
	XhoI- Yra1-fwd	CCCTCGAGTCTGCTAACTTAGATAAAATCC
	XhoI-dPCID-Yra1-fwd	AAActcGAGGTCAAGGTCAACGTCGA
	Apal-Yra1-rev	AAGGGCCCGCGATTTCAACAGTAAGA
	seqYra1Pro-fwd	CCTCGTTCTATTTGAAAC
	seqYra1ORF-fwd	CAAATCATGGCCTATAAGG
	seqYra1Term-rev	ATACATAATGAAACTCGG
Tho1-6xHA	Tho1-Xho_-500_fwd	agactcgagATACGTACGCTACGGTAAT
	Tho1-Sac1_+300_rev	agaggagctcCCACCTGCATTACTATGAAC
	Tho1_3'-EcoR1-rev	acagaattcTCTTCTGTAACCAGAGCG
	Tho1_3'-NheI-fwd	tgtgctagcAGGCTCAAGCCGTTTTG
	6xHA-pym15-EcoR1-fwd	tctgaattcCGTACGCTGCAGGTCGAC
	6xHA-pym15-NheI-rev	tgtgctagcTTAGCTAGAAGCGTAATCTG

Table 5: Oligo-nucleotides used for the introduction of affinity tags in this study

Name	DNA-sequence
Rpb3-TAP-ol1	ATGCATCTCAAATGGGTAATACTGGATCAGGAGGGTATGATAATGCTTGGtccatg gaaaagagaag
Rpb3-TAP-ol2	ATATAAAGCTTTTTTCTTATTATTTTCGGTTCGTTCACTTGTTTTTtacgactcact ataggg
Hpr1-TAP-ol1	ATGCAGCTACTTCGAACATTTCTAATGGTTCATCTACCCAAGATATGAAAtccatgg aaaagagaag
Hpr1-TAP-ol2	TAAAATCTATCTGAATTGTTGGGACACTATGCATGAATTTCTTATCAGTtacgactc actataggg
Tho2-TAP-ol1	TCCGCAAGGTCCCAAGGGTGGGAATTACGTCAGTAGGTACCAGAGGtccatggaaa agagaag
Tho2-TAP-ol2	TAGCTGGACCTTTTGATTTTACGAGATTACAATTCGGGAAGTATCtacgactcactat aggg
NTAP-Tho2-ol1	CGGTTTCAGTTGATACATATTCGCACCAGTATACATTTTCAGGACTTTgaacaaaagc tgagctcat
NTAP-Tho2-ol2	GAGAAAGAGCGTTCAATTTGGAAAGTAGCGTCTGTTCTGCCATcttatcgctcatcatca agtg
Mft1-TAP-ol1	GGCACTACAAGCGATTTTAGTGCGTCTTCTCTGTTGAAGAAGTAAAAtccatggaa aagagaag
Mft1-TAP-ol2	TTTCGGTTTCTATTGAACTGTAATAACTTGCCTATGTGTCTATATGCCcttatcgctcatc atcaagtg

Thp2-TAP-ol1	CATCGGAAAGCTATCCTGTAGATAAAGAAGGTGACATAGTTTTAGAAAtccatggaaa agagaag
Thp2-TAP-ol2	CAATTTTCTCTTTCTCCCTCCCACTCAACAAGTATGTATAACcttatcgatcatca gtg
Gbp2-TAP-ol1	CCTAAGCACCAACCGAACCATGACCTGTGTACTTTATGATAACGTATAAAAtccatgg aaaagagaag
Gbp2-TAP-ol2	AAATAAGGTTATTCTTTTGGTCAAAATGATAAATACTATTAAAGTTtacgactc actataggg
Hrb1-TAP-ol1	AACTATGGGGGTTGTGATTTGGATATATCGTACGCTAAACGCCTCtccatggaaaag agaag
Hrb1-TAP-ol2	TAAGTCAAGAATGAATAAATACTTGTGCGAGATCCAATAGGTGAtacgactcactata ggg
Nab2-TAP-ol1	CCGCAAACCAAGTTTTACGCACCAAGAACAAGATACGGAAATGAACtccatggaaaag agaag
Nab2-TAP-ol2	GGTGTCTTCATCAAAAGGGTCACAGGAACATGAATTCGTTCCGtacgactcactat aggg
NTAP-Npl3-ol1	TACTTTTGAAGGAATCAAAATTAAGCAATTACGCTAAAACCATAAGGATAgaa aagctggagctcat
NTAP-Npl3-ol2	ACAACAGATTCTGGTAGTTGCTCTACGTGAGTTTCTTGAGCTTCAGACATcttatcg catcatcaagtg
Spt5-TAP-ol1	CCAAGGAAATAAGTCAAACATGGTGGTAACAGTACATGGGGAGGTCAAtccatgg aaaagagaag
Spt5-TAP-ol2	CAGTATTTTCATCGAAACATCGAAATCTTCATAACTATTCTTATTAACGTGGtacgact cactataggg
Spt6-TAP-ol1	GACGCTTCTAAATCTAACAGTAGTAAGAATAGAATGAACAACTACCGTtccatgga aaagagaag
Spt6-TAP-ol2	CCTAAACAATGGTCAAAGTAATAATAAAATTAATAATAACAATGGACACtacgactc actataggg
Pcf11-TAP-ol1	CTAATAGTGGCAAGGTCGGTTTGGATGACTTAAAGAAATTGGTCACAAAAtccatg gaaaagagaag
Pcf11-TAP-ol2	CAAATGAGCAACCCATCAAATACCAAGAACAAGAATTTAATATAATATAtacga ctcactataggg
YCT-Ribo-ol1	GCACTTCTTGAAGGTTGAACAAAACATCAAGAGAAAATAAAAAAGAAAATATAA CATG CGTACGCTGCAGGTGCAC
YCT-Ribo-ol2	TCGAAGGAACTAGGATTTTCATCAGAAGAGGAGATCGAGTCTGCTC CCACTAGTGGATCTGATATC
Mft1-His6-ol1	GGCACTACAAGCGATTTTAGTGCCTCTTCTCTGTTGAAGAAGTAAACGTACGC TGCAGGTGCAC
Mft1-His6-ol2	TTTCGGTTTCTATTGAACTGTAATAACTTGCCTATGTGTCTATATGCCATCGATGA ATTCGAGCTCG
Rpc160-TAP-Col	GTCGATCGGTACCGGTTCC

Table 6: Real-time PCR primers used in this study

Name	DNA-sequence
qYER-fwd = NTR	TGCGTACAAAAAGTGTCAAGAGATT
qYER-rev = NTR	ATGCGCAAGAAGGTGCCTAT
qPMA1-P-fwd	GCTCCCTCCATTAGTTTCGA
qPMA1-P-rev	TGTTAGACGATAATGATAGGACATTTGAA
qPMA1-55-fwd	CAGCTCATCAGCCAACCTCAAGA
qPMA1-55-rev	AAGATTGAGATGCAGCGTCATC
qPMA1-5-fwd	CGATGACGCTGCATCTGAA
qPMA1-5-rev	CCGTGATTAGATTGTAGTTCTTCGATT

qPMA1-M5-fwd	CCTCTTCATCATCCTCTTCAGC
qPMA1-M5-rev	GCAGGCTTTTCTTGAGTTGG
qPMA1_M_fwd	AAATCTTGGGTGTTATGCCATGT
qPMA1_M_rev	CCAAGTGTCTAGCTTCGCTAACAG
qPMA1-M3-fwd	TTCGCTGATGTTGCTACTTTGG
qPMA1-M3-rev	TCCATTTAACGGGCTTTGGA
qPMA1-3-fwd	CGTCTTCGCTGTGACATCA
qPMA1-3-rev	TTTCAGACCACCAACCGAAT
qPMA1-Term-fwd	GGCATCAACTCTTAGCTTCACACA
qPMA1-Term-rev	CCATCCAGAGAAACCAATTATATCAA
qPMA1-pA1-fwd	TCAATGGGAACTGATACACTAAAAA
qPMA1-pA1-rev	AAGCGGCTTATTCTTGTTGG
qPMA1-pA2-fwd	ATAATGCACACACTAATTATTTCAT
qPMA1-pA2-rev	CCAAATATTTCAAATCCTACTCAA
qADH1-5-fwd	GGCCAACGAATTGTTGATCA
qADH1-5-rev	CGTGCCAAGCGTGCAA
qADH1-M-fwd	AGCCGCTCACATTCTCAAG
qADH1-M-rev	ACGGTGATACCAGCACACAAGA
qADH1-3-fwd	TTGGACTTCTCGCCAGAGG
qADH1-3-rev	GCCGACAACCTTGATTGGAG
qYEF3-5-fwd	GCCAAGGGTATCAAGGACAA
qYEF3-5-rev	GCAATGTGAGCAACAGCTTG
qYEF3-M-fwd	CATTGGTGCTGATTTGATCG
qYEF3-M-rev	TCATGTATGGGGTGATGTGG
qYEF3-3-fwd	CCGAATTCGTTAAGAAGTGTCC
qYEF3-3-rev	CCTGGTTCTGGGAACCTGAA
qGFP-5-fwd	GTGTTGTCCCAATTTGGTTGA
qGFP-5-rev	TTCACCGGAGACAGAAAATTTGT
qGFP-M-fwd (-500bp)	CCAGTTCCATGGCCAACCT
qGFP-M-rev (-500bp)	TGGGTATCTCGCAAAACATTGA
qGFP-3-fwd	CCAATTGGTGATGGTCCAGTCT
qGFP-3-rev	GTCTTTTCGTTTGGATCTTTGGAT
qYCT1-5-fwd (+100bp)	TCGCTGATGTTACGCTAGCATT
qYCT1-5-rev (+100bp)	CAGGCGTGATTTCTGGAACCTG
qYCT1-5M-fwd (+400bp)	TTTGTCTCTGCCATTAACCTTTTACTTT
qYCT1-5M-rev (+400bp)	GAAGCCTAGAATCGAATCATAGGATAA
qYCT1-M-fwd (+700bp)	TGAAGGAATGGGTGGAGTTGA
qYCT1-M-rev (+700bp)	CCAGCACCGGCAACTGAT
qYCT1-M3-fwd (+900bp)	ATCCCAGCATTTTCGTTATGGA
qYCT1-M3-rev (+900bp)	CAACAACCGTCATGATAAAAGTCAA
qYCT1-3-fwd (+1300bp)	AGTGCATCTGGGTACACCAAAA
qYCT1-3-rev (+1300bp)	CAATCCCATCGCAAATAAGG

1.6 peptides

Peptides were ordered from PSL (Heidelberg) or PANATECS (Tübingen).

Table 7: Sequences of the CTD peptides used in the pull-down experiments

Peptide	Sequence	Company
CTD	Biotin - YSPTSPS YSPTSPS YSPTSPS	PSL (Heidelberg)
S2P	Biotin - YSPTSPS <u>Yp</u> SPTSPS <u>Yp</u> SPTSPS	PSL (Heidelberg)
S5P	Biotin - YSPTSPS YSPT <u>p</u> SPS YSPT <u>p</u> SPS	PSL (Heidelberg)
Y1P	Biotin - YSPTSPS <u>pY</u> SPTSPS <u>pY</u> PTSPS	PSL (Heidelberg)
Y1P-S2P	Biotin - YSPTSPS <u>pYp</u> SPTSPS <u>pYp</u> SPTSPS	PANATECS (Tübingen)
Y1P-S5P	Biotin - YSPTSPS <u>pY</u> SPT <u>p</u> SPS <u>pY</u> SPT <u>p</u> SPS	PANATECS (Tübingen)
S2P-S5P	Biotin - YSPTSPS <u>Yp</u> SPT <u>p</u> SPS <u>Yp</u> SPT <u>p</u> SPS	PANATECS (Tübingen)

pY: phospho-tyrosine; pS: phospho-serin

1.7 Buffers, solutions and Growth media

Table 8: Buffers, solutions and media

Buffer/solution/media	Components
ChIP-elution	50 mM TRIS; 10 mM EDTA; 1% SDS; pH 7.5
Coomassie staining	0.25% (w/v) Coomassie Brilliant Blue R-250; 30% (v/v) ethanol; 10% (v/v) acetic acid
Coomassie destaining	30% (v/v) ethanol; 10% (v/v) acetic acid
DNA loading dye	6x 40% (w/v) sucrose; 0.25% Bromophenol Blue; 0.25% Xylene cyanol FF
FA-LS	50 mM HEPES; 150 mM NaCl; 1 mM EDTA; 1% Triton-X100; 0,1% SDS; 0,1% NaDeoxycholat; pH 7.5
FA-HS	50 mM HEPES; 500 mM NaCl; 1 mM EDTA; 1% Triton-X100; 0,1% SDS; 0,1% NaDeoxycholat; pH 7.5
KNOP-buffer 10x	50 mM TRIS-HCl, pH 9.2; 160 mM (NH ₄) ₂ SO ₄ ; 22.5 mM MgCl ₂
KNOP- Polymerase	2U Taq; 0.56U Vent
Luria-Bertani Broth (LB)	1% (w/v) tryptone; 0.5% (w/v) yeast extract; 0.5% (w/v) NaCl; (2% (w/v) agar for plates)
Phosphatase Inhibitors (PhosI) 100x	10 mM NaF; 1 mM Na ₃ VO ₄ ; 5 mM β-glycerophosphate
Phosphate-buffered saline (PBS)	137 mM NaCl; 2,7 mM KCl; 20 mM NaH ₂ PO ₄ ; 10 mM Na ₂ HPO ₄ , pH 7.5
PD-HS	1 M NaCl, 25 mM Tris/HCl, pH 8.0, 5% Glycerol, 2.5 mM DTT, 0.025% NP-40, 0.1% BSA, 1x PI, 1xPhosI
PD-LS	100 mM NaCl, 25 mM Tris/HCl, pH 8.0, 5% Glycerol, 2.5 mM DTT, 0.025% NP-40, 0.1% BSA, 1x PI, 1xPhosI
Protease Inhibitors (PI) 100x	8 ng/ml Leupeptin; 137 ng/ml Pepstatin A; 17 ng/ml PMSF; 0.33 mg/ml Benzamidine; in 100% EtOH p.a.
Sample buffer (SB)	4x 0.2 M TRIS-HCl, pH6.8; 40% (v/v) glycerol; 8% (w/v) SDS; few

	grains Bromophenol Blue; 0.1M DTT
SDS-PAGE Electrophoresis buffer	25 mM Tris; 0.1% (w/v) SDS; 0.19 mM glycine
Separation gel buffer 4x	3 M Tris; 0.4% (w/v) SDS; pH 8.8
Stacking gel buffer 4x	0.5 M Tris; 0.4% (w/v) SDS; pH 6.8
Solution I	10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 100 mM LiOAc
Solution II	10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 100 mM LiOAc; 40% (v/v) PEG-4000
Synthetic complete medium (SDC)	0.67% (w/v) yeast nitrogen base; 0.06% (w/v) complete synthetic mix of aa; drop out as required; 2% (w/v) glucose; when required 0.1% (w/v) 5-FOA was added; (2% (w/v) agar plates)
TAE 50x	2 M TRIS; 100 mM EDTA, pH 8.0; 1 M acetic acid
TE	1 mM EDTA; 10 mM TRIS-HCl, pH 8.0
TLEND	10 mM TRIS; 0.25 M LiCl; 1 mM EDTA; 0.5% NP-40; 0.5% NaDeoxycholat; pH 8.0
Tris-buffered saline (TBS)	137 mM NaCl; 2.7 mM KCl; 12.5 mM Tris-HCl
TSNTE	buffer 2% (v/v) Triton X-100; 1% (v/v) SDS; 100 mM NaCl; 10 mM TRIS-HCl, pH 8.0; 1 mM EDTA
Wet blotting buffer	25 mM TRIS; 192 mM glycine; 10% methanol
Yeast full medium (YPD)	2% (w/v) peptone; 2% (w/v) glucose; 1% (w/v) yeast extract; (2% (w/v) agar for plates)

1.8 Antibodies

Table 9: Antibodies used for Western-blotting

Name	Source	Dilution	Supplier
anti-myc	rabbit, polyclonal	1:5000	Upstate cell signaling solutions (#06-549)
anti-Yra1	rabbit, polyclonal	1:2000	(Strässer et al. 2002)
anti-Sub2	rabbit, polyclonal	1:10000	(Strässer et al. 2002)
Peroxidase Anti-Peroxidase	rabbit, monoclonal	1:5000	Sigma (P1291)
anti-HA	rat, monoclonal	1:1000	Roche (1-867-423)
anti-His	mouse, monoclonal	1:1000	ABM (G020)
anti-PGK	mouse, monoclonal	1:10000	Molecular probes (A6457)
anti-CBP	rabbit, polyclonal	1:2000	OpenBiosystems (CAB1001)
anti-rabbit-HPRO	goat, monoclonal	1:3000	Biorad (#170-6515)
anti-mouse-HPRO	goat, monoclonal	1:3000	Biorad (#170-6516)
anti-mouse-HPRO	goat, monoclonal	1:5000	Sigma (A9037)

Table 10: Antibodies used in ChIP experiments

Name	Source	Amount per ChIP	Supplier
anti-HA	rat, monoclonal	10µl	Roche (1-867-423)
anti-Rpb1 (Y-80)	rabbit, polyclonal	15µl	Santa Cruz Biotechnology (sc-25758)
anti-CTD-Tyr1	rat, monoclonal	50µl	3D12 (Mayer et al. 2012a)
anti-CTD-Ser2	rat, monoclonal	20µl	3E10 (Chapman et al. 2007)
anti-CTD-Ser5	rat, monoclonal	20µl	3E8 (Chapman et al. 2007)
anti-Yra1	rabbit, polyclonal	15µl	(Strässer et al. 2002)
anti-Sub2	rabbit, polyclonal	15µl	(Strässer et al. 2002)

2. Methods

2.1 General techniques

2.1.1 Molecular cloning

Molecular cloning was done according to Sambrook, J. and Russell, D. W. (2001). In this study restriction digests, ligations, transformation of vectors in *Escherichia coli* (DH5 α), and separation of DNA in agarose gels were used. Commercially available kits and enzymes were used according to manufacturer's instructions in the manual. Restriction enzymes, phosphatase (FastAP) and Ligation kits were purchased from Fermentas (St. Leon-Roth) and used as recommended by the manual. For DNA preparations from *E. Coli* the following kits were used: Nucleobond AX PC100 Macherey&Nagel (Düren), Nucleospin Mini Macherey&Nagel (Düren). Agarose Gel electrophoresis was performed in 1-2% Agarose gels buffered with 1x TAE at 180 V for 10-20 min, with GelRed™ stain (VWR, Ismaning). The Nucleospin extract Macherey&Nagel (Düren) Kit was used for extraction of DNA from PCR reactions or Agarose gels. Special cloning strategies are detailed below. All plasmids were verified by sequencing by MWG, Eurofins (Munich).

2.1.2 PCR

All PCRs for cloning were done in a 20 μ l Phusion Flash® High fidelity PCR reaction. PCRs for integrations into yeast were carried out using the KNOP Polymerase mix (see below) and the DNA from 300 μ l reaction was purified afterwards by Phenol/Chloroform extraction. One volume of phenol:chloroform:isoamylalcohol (25:24:1) was added and after 10 min centrifugation at 15000 g, the aqueous phase was extracted with an equal volume of chloroform. DNA was precipitated with 1/10 volume of 3 M NaOAc, pH 5.2 and 2 volumes of 100% ethanol at -20 °C. After washing with 70% ethanol, the DNA was dried and resuspended in 20 μ l TE buffer.

Phusion Flash® High fidelity PCR reaction and amplification cycle

0.5 μ l	template DNA (100-250 ng)	98 °C	1 min	35x
0.25 μ l	fwd primer (10 pmol/ml)	98 °C	30 sec	
0.25 μ l	rev primer (10 pmol/ml)	45-55 °C	30 sec	
10 μ l	2x Phusion PCR Master Mix	72 °C	15 sec/1 kb	
9 μ l	ddH ₂ O	72 °C	5 min	

KNOP PCR reaction and amplification cycle

1.0 µl	template DNA (100-250 ng)			
0.5 µl	fwd primer (100 pmol/ml)			
0.5 µl	rev primer (100 pmol/ml)	94 °C	2 min	35x
10 µl	10x KNOP PCR buffer	94 °C	1 min	
8 µl	dNTPs (25 mM each)	50 °C	30 sec	
2 µl	KNOP Polymerase Mix	68 °C	2:30 min	
78 µl	ddH ₂ O	68 °C	10 min	

2.1.3 Special cloning strategies**2.1.3.1 Cloning of the ribozyme constructs**

The ribozyme templates were synthesized by Thermo Fisher. Sequences for the active and inactive ribozymes itself were described in (Fong et al. 2009) and the used DNA sequences for cloning are detailed in Table 4. First the selection marker of pYM-N25 was replaced by the KanMX selection marker of pYM-N22 (Janke et al. 2004) using standard cloning techniques and the primers given in Table 4. The oligo nucleotide templates for the active and inactive hepatitis δ ribozyme were amplified using Phusion Flash® High fidelity PCR and cloned using EcoRI and EcoRV into the modified pYM-N25.

2.1.3.2 Cloning of Y1F CTD mutants

The CTD substitutions have been generated similar as in (West and Corden 1995). Oligo nucleotides were designed to encode for seven CTD repeats with or without substitutions as given in Table 4. Oligo nucleotides were annealed in annealing buffer (10 mM TRIS/HCl, 50 mM NaCl, 1 mM EDTA, pH 7.5) in 50 µl final volume with each 5 µl forward and reverse primer (100 pMol/ml), by heating for 5 min to 95 °C and removing the tube together with the metal incubator block and slowly cooling down to ambient temperature. pSB0 (West and Corden 1995) was digested using Aval and the annealed oligo nucleotides encoding seven CTD repeats were ligated into the plasmid. After checking insertion by colony PCR, a second round of digestion and ligation of an annealed repeat sequence was performed. The CTD fragment containing the desired 14 CTD repeats was subcloned by KpnI and SnaBI into pY1 to create the Rpb1-CTD mutant.

2.1.3.3 Cloning of Yra1-ΔPCID

The Yra1-ΔPCID mutant and the corresponding wildtype were cloned similar as described in (MacKellar and Greenleaf 2011). First the endogenous Yra1 promoter (-333 bp to -1 bp relative to the start codon of *YRA1*) was amplified from genomic DNA, inserting just downstream of the Start-codon a 3xHA tag with the reverse primer. The primers are given in Table 4. After inserting the Yra1 promoter by *SacI* and *XhoI* into pRS313, *YRA1* (encoding amino acids 1-226) and *yra1-ΔPCID* (encoding amino acids 77-226) were amplified including the *YRA1* terminator sequence (+ 621 bp) from genomic DNA and inserted using *XhoI* and *Apal* into the pRS313-ProYra1 construct. Both *YRA1* constructs were subcloned using *SacI* and *Apal* into pRS315. Expression of the construct in yeast was verified using Western blotting.

2.1.4 SDS PAGE

Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) was done according to (Laemmli 1970) using the Mini-Protean II system (Biorad). After electrophoresis, separated proteins were either transferred to a membrane (Western blot) or directly stained with Coomassie for 1 h. Gels were destained for 1 h with destaining solution and fixation was carried out using 10% acetic acid solution. To analyze proteins for their identity, Coomassie stained bands were cut out and send in 100 µl H₂O for mass spectrometry analysis to the Zentral Labor für Proteinanalytik (ZfP, lab of Prof. Axel Imhoff, Butenandt Institute, Munich).

2.1.5 Western-Blot

For Western blotting, proteins were transferred from and SDS-PAGE gel onto a nitro-cellulose membrane (Porablot NCL, Machery Nagel, Düren) using a wet blotting device (Biorad): 3 layers of Whatman paper followed by a nitro-cellulose membrane, the SDS-PAGE gel and 3 more layers of Whatman paper, all presoaked in wet blotting buffer, were assembled in the blotting device (Biorad). Proteins were transferred at 100 V for 75 min. Transfer was verified by staining with Ponceau S. Membrane was blocked with 2% milk powder in 1x PBS and incubated with the primary antibody at 4 °C overnight. Antibodies and dilutions are given in Table 9. Membrane was washed 3x with 2% milk powder in 1x PBS at RT for 10 min and then incubated with the corresponding secondary antibody at RT for at least 1 h. The immunostained proteins were detected with ECL kit (Applichem). After

exposure of the membrane to the light sensitive films, the films were developed with a Kodak X omat M35 developing machine.

2.2 Yeast culture

2.2.1 Culture of *Saccharomyces cerevisiae*

Yeast strains were cultured in either full media or synthetic complete (SC) medium, lacking one specific amino acid or nucleotide to force the cells to keep the transformed plasmids. Cell density of the yeast culture was determined in a spectrophotometer at a wavelength of 600 nm. One optical density unit (OD) at 600 nm corresponds to *app.* 2.5×10^7 cells.

2.2.2 Transformation of yeast cells

50 ml of a yeast culture were grown to log phase (0.6-0.8 OD₆₀₀) and harvested by centrifugation (3600 rpm, 3 min). Cells were washed with 10 ml of H₂O and with 500 µl of solution I and then resuspended in 200 µl solution I. 1-5 µg of DNA were mixed with 5 µl single stranded carrier DNA (DNA of salmon or herring testis, 2 mg/ml). 50 µl cell suspension and 300 µl of solution II were added. After 30 min incubation at room temperature, cells were heat shocked (42 °C, 10 min) and incubated 3 min on ice. After addition of 1 ml H₂O, cells were pelleted. When plasmids were transformed, yeast was immediately plated on the selective plates in 100 µl H₂O. For genomic integration, cells were resuspended in 1 ml of YPD and incubated 1-5 h at room temperature on a turning wheel and then plated on selective media.

2.2.3 Whole cell extracts (WCE)

For testing integration of tags, WCE were prepared from yeast restreaked and freshly grown on plates: one white inoculation loop of cells was resuspended in 75 µl of 1x SDS-sample buffer and *app.* 50 µl of glass beads were added. Samples were incubated 45 sec at 95 °C, and vortexed for 45 sec. This was repeated once and then samples were again incubated 1 min at 95 °C.

For protein quantification, denaturing WCE were prepared as in (Knop et al. 1996): 5 OD₆₀₀ of logarithmically growing cells were washed once with 5 ml H₂O, resuspended in 500 µl of H₂O, 150 µl of pre-treatment solution (1.85 M NaOH, 7.5% β-mercaptoethanol) was added. After 20 min incubation on ice, 150 µl of 55% TCA were added to precipitate proteins. After 30 min centrifugation

at 15000 g, 4 °C, the protein pellets were resuspended in 100 µl of 1x SDS-sample buffer and 10 µl of 1 M TRIS base.

2.3 TAP-Purification

Tandem affinity purifications (TAP) were performed essentially as described in (Strässer et al. 2002). For the preparation of protein complexes from yeast 2 l of OD₆₀₀ 3-3.6 yeast grown in YPD were lysed using glass beads in a planetary mill (Pulverisette, Fritsch). The lysate was cleared from cellular debris for 10 min by centrifugation at 4000 rpm and 60min at 100000 g at 4 °C. The cleared lysate was bound for 1 h at 4 °C to 400 µl Fast Flow IgG Sepharose beads (GE). After washing with TAP buffer according to Table 11, the proteins were eluted using 150 µl 100 mM NaCl TAP-buffer and 3 µl TEV protease by incubation for 75 min at 16 °C.

For the Calmodulin purification step, the TEV-eluate was bound to 500 µl prewashed Calmodulin Sepharose beads (Stratagene) for 1 h at 4 °C in TAP buffer supplemented with 2 mM CaCl₂. After washing the beads in Mobicol with 10 ml 100 mM NaCl TAP buffer with 2 mM CaCl₂, the proteins were eluted in elution buffer (5 mM EGTA, 50 mM TRIS/HCl pH 7.5) at 37 °C for 15 min. Proteins were precipitated with 10% TCA for 20 min on ice and 30 min centrifugation at 4 °C with 13000 rpm. Protein pellets were resuspended in 60 µl 1x SDS-sample buffer and pH was neutralized using 1 M TRIS base.

Table 11: Buffer conditions for Tandem-Affinity-Purifications

Condition	Buffer	Proteins/ complexes
Low salt	- 2x 10 ml: 100 mM NaCl, 50 mM TRIS/HCl, pH 7.5, 1.5 mM MgCl ₂ , 0.15% NP-40, 1 mM DTT, 1x PI	TREX (Hpr1-TAP, Tho2-TAP, TAP-Tho2)
Medium salt	- 2x10 ml: 250 mM NaCl, 50 mM TRIS/HCl, pH 7.5, 1.5 mM MgCl ₂ , 0.15% NP-40, 1 mM DTT, 1x PI - 2 ml: 100 mM NaCl, 50 mM TRIS/HCl, pH 7.5, 1.5 mM MgCl ₂ , 0.15% NP-40, 1 mM DTT, 1x PI	Pcf11-TAP, Rix1-TAP
High salt	- 2x 10 ml: 1000 mM NaCl, 50 mM TRIS/HCl, pH 7.5, 1.5 mM MgCl ₂ , 0.15% NP-40, 1 mM DTT, 1x PI - 2 ml: 100 mM NaCl, 50 mM TRIS/HCl, pH 7.5, 1.5 mM MgCl ₂ , 0.15% NP-40, 1 mM DTT, 1x PI	THO (Hpr1-TAP)

To yield highly pure THO or TREX complex for peptide pull down assays, TEV eluates of the Hpr1-TAP, Mft1-His₇-Myc strain were bound to 250 µl Ni-NTA agarose (Qiagen) in the according TAP buffer, which was supplemented with 20 mM imidazole for 1 h at 4 °C, washed with 10 ml of the according TAP buffer with 20 mM imidazole and eluted in 2x 75 µl elution buffer (100 mM NaCl TAP buffer, 250 mM imidazole).

2.4 CTD-Peptide Pull-down Assay

Pull-down assays were performed as described in (Qiu et al. 2012) with small modifications: For each pull-down assay 15 µl of Streptavidin coupled magnetic beads M280 (Invitrogen) were washed three times with PD-HS buffer. Beads were resuspended in 100 µl HS buffer and incubated with 10 µg of each peptide for 2 h at 4°C. Peptides sequences are listed in Table 7. Peptides were ordered from PSL (Heidelberg) and PANATECS (Tübingen). Subsequently, beads were washed once with PD-HS buffer and two times with PD-LS buffer. For alkaline phosphatase (AP) treatment samples were washed two times with 1x fast digestion buffer (Fermentas), incubated for 15 min at 37 °C with 25 U FastAP (Fermentas), washed 2x with PD-LS buffer and resuspended in 100 µl PD-LS buffer. To test CTD binding, equal amounts of the different protein complexes (typically 5-10 µl) were incubated with the CTD-coupled beads for 90 min at 4°C. The non-bound fraction was collected. After 4 washing steps with 500 µl PD-LS buffer beads were resuspended in 1x SDS-sample buffer to elute the bound protein complexes. Non-bound and bound protein complexes were detected with an anti-CBP antibody (Open Biosystems) recognizing the remaining CBP-tag on the tagged proteins (Hpr1, Rix1 and Pcf11, respectively).

2.5 Small scale ChIP

ChIP was essentially performed as described in (Rother et al. 2010). Briefly, yeast was grown in the appropriate medium from OD₆₀₀ 0.1 - 0.2 to 0.8 and cross-linked with 1% formaldehyde for 20 min at 20 °C. After stopping the crosslinking reaction by addition of 0.25 M glycine and incubation for 5 min at 20 °C, cells were harvested by centrifugation for 5 min at 4 °C, 4000 rpm and washed three times with 1x TBS buffer and frozen in liquid nitrogen. Cells were lysed in FA-LS buffer with equal volumes of glass beads for 5 times 3 min with 3 min breaks on ice by vortexing at the highest energy setting. After sonication using a Bioruptor for 3 times 15 min, which yielded an average chromatin size of 250 bp, the lysate was cleared by 5 min and then 15 min centrifugation at 13000 rpm. 15 µl of the lysate were saved as input sample. The remaining lysate was incubated with 15 µl IgG coupled magnetic

beads (Invitrogen) for TAP-tagged proteins. For ChIP using antibodies, lysates were incubated for 2 h (20 °C) with 50 µl αY1P, 20 µl αS2P, αRpb1 (Y-80, Santa Cruz Biotechnology), αYra1 (Strässer et al. 2002) or αSub2 (Strässer et al. 2002) antibodies and 1.5 h (20°C) with 15 µl Protein A or G coupled magnetic beads (Invitrogen). Afterwards beads were washed with each 1 ml of 2x FA-LS, 2x FA-HS, 1x TLEND and 1x TE buffer. DNA was eluted for 30 min at 65°C under shaking in 140 µl ChIP elution buffer. 80 µl of 1x TE and 10 µl ProteinaseK (10 mg/ml, Sigma) were added to the supernatant and incubated 2 h at 37 °C and 12-16 h at 65 °C. Afterwards, DNA was purified using a PCR clean up Kit (Machery-Nagel).

The Avitag ChIP was performed similar as described above with modifications according to van Werven and Timmers (2006): Yeast strains were grown in SDC-Ura supplemented with 250 nM Biotin. After cell lysis, insoluble chromatin was pelleted (13000 rpm, 4 °C, 15 min), resuspended in 1 ml of FA-LS buffer, pelleted and resuspended again, to deplete the lysate from cytoplasmic, biotinylated proteins. IP was carried out with 50 µl Streptavidin magnetic beads M280 (Invitrogen) for 2 h 30 min at 4 °C. Beads were washed 2x with buffer A (10 mM TRIS/HCl, 0.5 M LiCl, 1 mM EDTA, 0.5% NP40, 0.5% NaDeoxycholat, pH 8.0), 3x with buffer B (10 mM TRIS/HCl, 1 mM EDTA, 3% SDS, pH 8.0), two times with TLEND buffer and once with 1x TE. DNA was eluted for 16 h at 65 °C with 1000 rpm shaking in 140 µl of Avi-elution buffer (1% SDS, 500 mM NaCl, 50 mM TRIS/HCl, pH 8.0) and 1 µl RNaseA (10 mg/ml, Fermentas). Afterwards, proteins were digested with 20 µl Proteinase K (10 mg/ml, Sigma Aldrich) for 2 h at 37 °C. DNA was purified as above.

2.6 qPCR

The DNA from the ChIP experiments was analyzed by quantitative real-time PCR with a StepOnePlus™ Real Time PCR System (Applied Biosystems). Reaction mixtures (10 µl final volume) contained 2.5 µl of 1:10 dilutions of the individual DNA samples, 5 pmol forward and reverse primers and 5 µl Power SYBR® Green PCR Master Mix (Applied Biosystems) as described by the manual. A non-transcribed region (NTR) on Chromosome V (174131-174200) (YER) served to normalize for background. Primers were designed using Primer express 3.0 software and tested for specificity and dimer formation by melting curve analysis of the individual amplification products. Oligo nucleotide sequences of the primers are given in Table 6. The thermocycling program included an initial denaturation for 10 min at 95 °C, followed by 45 cycles of amplification, comprising denaturation at 95 °C for 15 s and annealing/elongation at 60 °C for 1 min. For melting curve analysis of the amplification products, the temperature was increased in 0.3 °C increments to a final temperature of 95 °C under continuous fluorescence measurement. All reactions were run in triplicates and included

negative controls (H₂O). Standard curves were used to estimate primer efficiencies and factor occupancies were calculated according to the comparative Ct method, relative to the non-transcribed region according to $(E^{(C_T^{IP}-C_T^{INP})})_{NTR}/(E^{(C_T^{IP}-C_T^{INP})})$ (Livak and Schmittgen 2001).

2.7 ChIP-chip

2.7.1 ChIP

ChIP-chip assays were performed similar as in (Mayer et al. 2010). ChIP-chip of Rpb3-TAP, TAP-Tho2, Tho2-TAP, Hpr1-TAP, Mft1-TAP, Gbp2-TAP, Hrb1-TAP, Nab2-TAP and TAP-Npl3 was performed using IgG coupled Fast Flow Sepharose beads (GE). Antibodies directed against the protein were used for Yra1, Sub2, Y1P, S2P and S5P.

For each ChIP-chip experiment 600 ml of yeast were grown from OD₆₀₀ 0.1 to 0.8 and cross-linked at 20 °C with 1% formaldehyde for 20 min at 20 °C. Crosslinking reaction was stopped using 0.25 M glycine for 15 min. Cells were harvested by centrifugation (5 min, 4 °C 4000 rpm), washed two times with 1x TBS, once with FA-LS buffer and flash frozen in liquid nitrogen. Cells were resuspended in 3x 2 ml FA-LS buffer, divided into 4x 2 ml tubes with each 800 µl glass beads and mixed for 5 times 3 min with 3 min breaks on ice. Chromatin was pelleted two times (13000 rpm, 4 °C, 15 min) and resuspended in 1 ml of FA-LS buffer. Chromatin was sheared using a Bioruptor UCD-200 (Diagenode) for 3x 15 min with 5 min cooling with ice in between, to yield a Chromatin fragments with an average of around 250 bp. Debris was pelleted by centrifugation (5 min, 13000 rpm, 4 °C and 15 min, 13000 rpm, 4 °C). Lysate was incubated for 4 h at 4 °C with 45 µl of IgG-Sepharose-beads (GE). Beads were washed with each 400 µl: 2x FA-LS buffer, 3x FA-HS buffer, 2x TLEND buffer and once with 1x TE. Elution was carried out for 1 h at 65 °C under shaking in 140 µl ChIP elution buffer. Beads were pelleted and supernatant was added to 80 µl TE and incubated 2 h at 37 °C with 20 µl Proteinase K (10 mg/ml Sigma) and at 65 °C for additional 16 h. DNA was purified using a QIAquick PCR purification Kit (Qiagen) as described in the manual. DNA was eluted in 50 µl H₂O. RNA was digested with 5 µl RNaseA (10 mg/ml Fermentas) for 20 min at 37 °C. DNA was afterwards purified again using the QIAquick PCR purification Kit (Qiagen). DNA was eluted in 50 µl H₂O. Volume was narrowed down in a DNA Speedvac (Thermo) to 10 µl.

2.7.2 Whole genome amplification (WGA)

To prepare DNA for the hybridization, a library preparation and whole genome amplification was performed using the GenomePlex© whole genome amplification Kit 2 (WGA 2, Sigma Aldrich) as

described in the manual and by the Farnham lab (O'Geen et al. 2006). Briefly 10 µl DNA from the ChIP was used for the generation of the PCR-amplifiable OmniPlex® Library as detailed in the WGA2 manual. The DNA fragments in this library have constant flanking regions which can be used to anneal primers for a subsequent PCR. After a first round of amplification, which was performed exactly as described in the manual, the resulting DNA was purified using the QIAquick PCR purification Kit (Qiagen) and eluted with 50 µl ddH₂O. DNA concentration was determined using a ND-1000 Spectrophotometer (NanoDrop, Thermo Fisher). 150 ng of DNA were amplified in a second round of whole genome amplification. The reaction was carried out as described by the WGA 2 manual, but with the addition of 0.4 mM dUTP. The dUTP can be used for enzymatic fragmentation of the amplified DNA to smaller pieces which can be hybridized to the array. This is described in the following section. The DNA was purified using the QIAquick PCR purification Kit (Qiagen) and eluted with 50 µl ddH₂O. DNA quality was again assessed using a ND-1000 Spectrophotometer (NanoDrop, Thermo Fisher) and with agarose gel electrophoresis.

2.7.3 Array hybridization

Preparation of DNA and hybridization was performed as described by Affymetrix (Affymetrix Chromatin Immunoprecipitation Assay Protocol P/N 702238). Enzymatic fragmentation of the DNA and terminal labeling was performed using the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (P/N 900812, Affymetrix). Briefly, 5.5 µg DNA yielded from the second whole genome amplification with incorporated dUTP were used for hybridization. The DNA was fragmented with 1.5 µl Uracil-DNA-glycosylase (10 U/µl) and 2.25 µl *ApeI* (100 U/µl) at 30 °C for 1 h 15 min. The average fragment size (50-70 bases) was checked by agarose gel electrophoresis (2% AGE). The DNA fragments were labeled on their 3' end using 2 µl Terminal nucleotidyl transferase (TdT, 30 U/µl) and 1 µl GeneChip DNA Labeling Reagent (5mM). Labeled DNA was hybridized to a high-density custom-made Affymetrix tiling array (David et al. 2006) (PN 520055) at 45 °C for 16 h with constant rotational mixing at 60 rpm in a GeneChip Hybridization Oven 640 (Affymetrix, Santa Clara, CA). Washing and staining of the tiling arrays were performed using the FS450_0001 script of the Affymetrix GeneChip Fluidics Station 450. The arrays were scanned using an Affymetrix GeneChip Scanner 3000 7G. The resulting .CEL files were used for the data analysis. The raw ChIP-chip data (.CEL files) have been submitted to Array Express with accession numbers (E-MTAB-1400).

2.8 Data analysis of ChIP-chip Data

2.8.1 Normalization of ChIP-chip data

The data normalization was carried out similar as in (Mayer et al. 2010) and (Mayer et al. 2012a). At least two independent, biological replicates were used for each protein, dependent on the correlation of the raw data signal correlation between the replicates. For quality control, correlations of raw data and normalized data have been tested. The whole data normalization procedure was carried out using R (RDevelopmentCoreTeam 2009) and Bioconductor (Gentleman et al. 2004). The Affymetrix .CEL files were imported into R by conversion into an ExpressionSet R-object, using the R package Starr (Zacher et al. 2010). Data normalization consisted of four steps: In the first step we determined outlierprobes, which were determined by kernel density estimation. Single signals lying in the lowest density, not correlating well between the single replicates, were omitted and interpolated by the neighboring, overlapping probes. In the second step we performed quantile normalization between the replicate measurements. In the third step we averaged the replicate measurements (including the reference measurements) by calculation of the geometrical average over each probe intensity for the replicates. In the fourth step, data from all factors were normalized using a combined mock IP (referring to ChIP-chip experiments with nontagged wild-type strain) plus input reference normalization. The normalized signal was thereby converted into occupancy values between 0-100 % by setting the genome-wide 99.8 % quantile to 100 % occupancy and the 10 % quantile to 0 % occupancy. More details of this normalization are given in (Mayer et al. 2010).

4.8.2 Gene occupancy profiles

To generate the occupancy profiles for each gene, the signals of all probes overlapping each nucleotide at the given position were used to calculate the median signal, which reflects the protein occupancy at this position. In average 6.5 probes per nucleotide were used, if not single outlier probes were omitted as described above. To correct for background noise in the data, the intensities were smoothed by using a sliding window smoothing procedure (window half size of 75 bp) as implemented in the R package Ringo (Toedling et al. 2007).

2.8.3 Meta profiles of gene classes

The gene classes used here were previously described in (Mayer et al. 2010). Only 'verified' or 'uncharacterized' genes annotated in the Saccharomyces Genome Database (5769 genes

<http://www.yeastgenome.org/>) were considered. Only genes with annotated Transcription Start Site (TSS) and polyA sites (pAs) were used (4366 genes), which were assigned by RNA-Seq (Nagalakshmi et al. 2008a). To remove possible wrongly annotated TSSs and pAs, we only included genes with TSS annotations showing a distance less than 200 bp to the corresponding downstream ATG codon or with a polyA site less than 200 bp downstream of the stop codon (3448 genes).

Since the yeast genome is built very compact and ChIP-chip usually has a limited resolution (median inter-ORF length: 368 bp, median inter-transcript length: 259 bp), a gene's factor occupancy profile can have spurious contributions from flanking genes. To avoid effects by "spill-over" from one gene to the next, we only used genes with a minimum ORF and transcript distance to neighboring genes of 250 bp and 200 bp, respectively (1786 genes). We restricted the analysis in the Meta profiles to the top 50% expressed genes, according to the measurements by (Dengl et al. 2009) (1140 genes). We grouped the genes into short genes (S) 512 to 937 bp, medium genes (M) 938 to 1537 bp, and long genes (L) 1538 to 2895 bp, comprising 266, 339, and 299 genes, respectively. The Meta profiles in each group were scaled to the median length and the Meta profiles were calculated by taking the median over the scaled profiles.

2.8.4 Length dependency plots

To analyze for the length dependency, profiles for all 'verified' and 'uncharacterized' genes were calculated as above. Genes were selected by being at least 250 bp upstream and downstream separated from other genes, according to the *SGD* database. For each gene the peak occupancy was calculated, reflecting the 90 % quantile of the factor recruitment. This gives a robust measurement for how much the highest occupancy of a protein is per gene. The peak occupancies of each factor were normalized to the peak occupancies of the RNAPII (Rpb3) signal for each gene to normalize for the transcription frequency. Genes were sorted into eight length groups: A (512-723 bp), B (724-1023 bp), C (1024-1286 bp), D (1287-1617 bp), E (1618-2047 bp), F (2048-2895 bp), G (2896-4095 bp) and H (4096-5793 bp). Mean values and standard deviations were calculated and plotted versus the gene length.

2.8.5 Correlations of recruitment

To analyze for the correlation of the different factors, peak occupancies were calculated as above. For each gene the peak occupancies were calculated, reflecting the 90% quantiles of the factor

recruitment to the genes. The Pearson correlation coefficients between the factors were calculated and the single peak occupancies were plotted as dots.

For the analysis of Nab2, RNAPII (Rpb3) and RNAPIII (Rpc160), protein coding genes which lie nearby (<251 bp) tRNA genes were stained in the plot, because the signal of the tRNA genes is likely to “spill over” to the protein coding genes, since tRNA genes are very heavily transcribed.

2.9 Expression profiling of Tho2-TAP

Experiments were carried out in collaboration with Eoghan O’Duibhir and Frank CP Holstege (UMC Utrecht, Molecular Cancer Research Institute in Utrecht, the Netherlands) as described in (Lenstra et al. 2011). Briefly, *THO2-TAP* and wild type cells were grown in SDC media with 2% glucose to mid log phase. Each strain was measured four times independently. RNA was isolated using robotic liquid handler and hybridized in dye-swap biological replicate to dual-channel 70-mer oligonucleotide arrays to obtain four independent measurements as described in (Lenstra et al. 2011). After quality control, normalization and dye-bias correction, Up- or down-regulation of expression in the *THO2-TAP* strain was defined as a > 1.7-fold change versus the average wild-type with a p-value of < 0.05. Further analysis was carried out in the Sträßer lab, using R (RDevelopmentCoreTeam 2009). The average length, GC-content, expression level, RNAPII occupancy, convergence, divergence, +1 and +2 nucleosome positioning of the up- and down-regulated genes was calculated. The statistical significance of each group was determined using the Wilcoxon rank sum test. SAGA and TFIID dominated genes were analyzed for expression changes versus all promoters using the Wilcoxon rank sum test. The expression data has been submitted to Array Express.

Abbreviations

app.	approximately
bp	basepair
ChIP	Chromatin Immunoprecipitation
CTD	C-terminal domain
CTR	C-terminal region
ds	double stranded
HA	Haemagglutinin
μ	micro
mRNA	messenger Ribonucleic acid
mRNP	messenger Ribonucleoprotein
NPC	nuclear pore complex
nt	nucleotide
nts	nucleotides
p	phospho
polyA	poly-Adenosine
Prp19C	Prp19 complex
qPCR	quantitative PCR
RNAPI, II, III	RNA Polymerase I, II, III
RRM	RNA recognition motif
RS	arginine and serine rich
S, Ser	serine
Ser2-Ser5	CTD repeat with two phosphorylations, one on Ser2 and Ser5
ss	single stranded
SR	splicing related
TAP	Tandem Affinity Purification
TREX	Transcription and export
THO	suppressor of transcriptional defect of Hpr1 by overexpression
Tyr	tyrosine
<i>wt</i>	wildtype
Y	tyrosine

Acknowledgements

I would like to thank...

... first of all, Katja Sträßer for giving me the opportunity to work on this challenging and exciting project. I am grateful for the scientific discussions, the scientific exchange and the constant support as well as believing in the project. I also would like to thank especially for the opportunity to follow up my own ideas and the encouraging in starting new projects.

... Johannes Soeding and Matthias Siebert for supervising and training me in data analysis and of course for always helpful discussions and ideas. This made the project possible.

... the members of my Thesis Advisory Comitee: Patrick Cramer, Johannes Söding and Elena Conti for constructive and fruitfull discussion. Especially I am grateful to Patrick Cramer for support of my PhD project and for being the second examiner of my thesis.

... IMPRS-LS for support, scientific exchange, excellent workshops and seminars as well as for having a good time.

... Cornelia Burkert-Kautzsch for taking part in the sometimes frustrating ChIP-chip experiments and sharing the ChIP-chip data with me.

... our collaborators Frank Holstege and Eoghan O'Duibhir for performing the expression arrays, Elisabeth Tran and Wai Kit Ma for Helicase assays and Luca Schenk and Andre Gerber for collaboration on Slf1.

... Andreas Mayer and Michael Lidschreiber for discussion and help with the ChIP-chip protocol, this was always a great help!

... Anja Kieser for excellent support, which really took this project to where it is and also for always enjoying working together with you. It is really great to have you in the team!

... Jan Strauß for always waiting patiently for Anja when experiments took longer.

... Max Reuter, for help and discussion and for proofing our hypothesis that Nab2 is indeed involved in RNAPIII transcription.

... the whole Sträßer lab, Anja, Connie, Britta, Chiara, Max, Rashmi, Dimtry, Viter, Shivendra, Sittinan, Benno, Juliane and Katharina for a allways nice atmosphere and help when ever needed.

... to my students Alex, Jessica, Mirijana and Verena for giving me the opportunity to teach and for the great help they have been.

Ganz besonders danken möchte ich:

... meiner Familie, dabei ganz besonders meinen Eltern Erich und Inge Meinel, für ihre Unterstützung, Aufmunterung, Hilfe und Geduld

... meiner Freundin Sabine Ströbl für ihre Unterstützung, ihre Hilfe, Aufmunterung, Beratung und vor allem für ihre Geduld. Es ist schön dass es Dich gibt!

Tables

Table 1: Equipment used in this study	68
Table 2: Yeast strains used in this study.....	69
Table 3: Plasmids used in this study.....	71
Table 4: Oligo-nucleotides used for cloning.....	72
Table 5: Oligo-nucleotides used for the introduction of affinity tags in this study.....	73
Table 6: Real-time PCR primers used in this study.....	74
Table 7: Sequences of the CTD peptides used in the pull-down experiments.....	76
Table 8: Buffers, solutions and media	76
Table 9: Antibodies used for Western-blotting.....	77
Table 10: Antibodies used in ChIP experiments.....	78
Table 11: Buffer conditions for Tandem-Affinity-Purifications	83

Figures

Figure 1: early mRNP formation.....	9
Figure 2: The yeast mRNA export pathway.....	11
Figure 3: The Rpb1-CTD.....	15
Figure 4: Phosphorylation of the Rpb1-CTD during the transcription cycle	17
Figure 5: mRNA export through the Nuclear Pore complex.	20
Figure 6: TREX recruitment increases from the 5' to the 3' end of the gene	23
Figure 7: Meta profiles of Hrb1, Gbp2, Nab2 and Npl3 are distinct different from THO, Sub2 and Yra1.	24
Figure 8: Gbp2-TAP and TAP-Gbp2 occupancy is indistinguishable from each other.....	25
Figure 9: Rpb3, Hpr1, Yra1 and Sub2 are recruited to protein coding and only in minor amounts to sn/snoRNA genes	26
Figure 10: Nab2 is recruited to RNAPIII transcribed loci and correlates with RNAPIII at tRNA loci.....	28
Figure 11: Nab2 binds <i>in vivo</i> to immature RNAPIII transcripts.....	29
Figure 12: Yra1 binds preferentially to non-intron containing genes, Npl3 is recruited to higher levels to intron containing genes.	30
Figure 13: TREX is recruited length dependent to the genes.....	31
Figure 14: THO/TREX occupancy increases throughout the ORF.....	32
Figure 15: Pearson Correlation coefficients between TREX components, Nab2, Npl3 and general elongation factors.	32
Figure 16: <i>bona fide</i> general elongation factors are not length dependent.....	33
Figure 17: Ribozyme assay scheme.	35
Figure 18: TREX, Nab2 and Npl3 depend on the nascent mRNA for efficient recruitment or occupancy.	35
Figure 19: RNA dependent occupancy cannot be detected over more than 400-700 bp	36
Figure 20: Tyr1 and Ser2 phosphorylations are length dependent and correlate with THO, Yra1 and Sub2.....	37
Figure 21: Proper Tyr1 and Ser2 CTD phosphorylation is essential for efficient THO recruitment.	38
Figure 22: Proper Ser2 CTD phosphorylation is needed for efficient TREX recruitment.	39
Figure 23: Ser2P levels are lower on sn/snoRNA genes, while Tyr1P is high.....	39
Figure 24: THO/TREX recruitment is independent of Yra1.	41
Figure 25: Yra1- Δ PCID does not assemble into the TREX complex.....	42
Figure 26: The THO complex can be purified without Yra1 using high salt conditions.....	43
Figure 27: THO binds directly to Ser2-Ser5-diphosphorylated CTD-peptides.....	44
Figure 28: <i>THO2-TAP</i> impairs recruitment of THO/TREX	46
Figure 29: Tho2-TAP does not impair TREX assembly.....	47
Figure 30: <i>THO2-TAP</i> impairs correct expression of long genes.	48
Figure 31: <i>THO2-TAP</i> is synthetic lethal with <i>yra1-ΔPCID</i>	49
Figure 32: Tho1 is recruited during elongation phase and Tho1 recruitment is length dependent.....	50
Figure 33: Tho1 occupancy correlates very well with THO/TREX	50
Figure 34: Tho1 recruitment to the transcription site is promoted by the nascent RNA	51
Figure 35: Δ <i>hpr1</i> abolishes Tho1 recruitment.....	52
Figure 36: Spt5-CTR S1A mutations lead to increased THO/TREX and decreases Tho1 occupancies... 53	
Figure 37: Spt5-CTR S1D phospho-mimicry mutants restore Tho1 levels and decrease THO/TREX levels.....	54

Figure 38: Overexpression of Tho1 leads to decreased THO and Sub2 levels.	55
Figure 39: Model of TREX recruitment and mRNP formation.....	61
Figure 40: Recruitment profiles of RNAPII, THO/TREX and Pcf11	62
Figure 41: predicted domain architecture of Gbp2, Hrb1 and Npl3	65

References

- Abruzzi KC, Lacadie S, Rosbash M. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J.* 2004; **23** (13): 2620–2631.
- Alcázar-Román AR, Tran EJ, Guo S, Wentz SR. Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. *Nat Cell Biol* 2006; **8** (7): 711–716.
- Anderson JT, Wilson SM, Datar KV, Swanson MS. NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol. Cell. Biol.* 1993; **13** (5): 2730–2741.
- Ares M, JR, Grate L, Pauling MH. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* 1999; **5** (9): 1138–1139.
- Barilla D, Lee BA, Proudfoot NJ. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences* 2001; **98** (2): 445–450.
- Bataille AR, Jeronimo C, Jacques P, Laramée L, Fortin M, Forest A, Bergeron M et al. A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Molecular Cell* 2012; **45** (2): 158–170.
- Bourbon H. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Research* 2008; **36** (12): 3993–4008.
- Bucheli ME, Buratowski S. Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. *EMBO J* 2005; **24** (12): 2150–2160.
- Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadgil O, Lesne A, Buc H et al. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* 2006; **441** (7094): 770–773.
- Carty SM, Goldstrohm AC, Sune C, Garcia-Blanco MA, Greenleaf AL. Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proceedings of the National Academy of Sciences* 2000; **97** (16): 9015–9020.
- Chanarat S, Burkert-Kautzsch C, Meinel DM, Sträßer K. Prp19C and TREX: Interacting to promote transcription elongation and mRNA export. *transcription* 2012; **3** (1): 8–12.
- Chanarat S, Seizl M, Strasser K. The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes & Development* 2011; **25** (11): 1147–1158.
- Chang Y, Imam JS, Wilkinson MF. The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu. Rev. Biochem.* 2007; **76** (1): 51–74.
- Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E et al. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* 2007; **318** (5857): 1780–1782.
- Chávez S, Beilharz T, Rondón AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T et al. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *EMBO J.* 2000; **19** (21): 5824–5834.
- Chavez S, Garcia-Rubio M, Prado F, Aguilera A. Hpr1 Is Preferentially Required for Transcription of Either Long or G+C-Rich DNA Sequences in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 2001; **21** (20): 7054–7064.

- Cheng H, Dufu K, Lee C, Hsu JL, Dias A, Reed R. Human mRNA Export Machinery Recruited to the 5' End of mRNA. *Cell* 2006; **127** (7): 1389–1400.
- Chen YG, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Research* 2007; **35** (12): 3928–3944.
- Cho E, Kobor MS, Kim M, Greenblatt J, Buratowski S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & Development* 2001; **15** (24): 3319–3329.
- Cho E, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & Development* 1997; **11** (24): 3319–3326.
- Cramer P. RNA polymerase II structure: from core to functional complexes. *Current Opinion in Genetics & Development* 2004; **14** (2): 218–226.
- David CJ, Boyne AR, Millhouse SR, Manley JL. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes & Development* 2011; **25** (9): 972–983.
- David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T et al. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U.S.A.* 2006; **103** (14): 5320–5325.
- Deka P, Bucheli ME, Moore C, Buratowski S, Varani G. Structure of the yeast SR protein Npl3 and Interaction with mRNA 3'-end processing signals. *J. Mol. Biol.* 2008; **375** (1): 136–150.
- Dengl S, Mayer A, Sun M, Cramer P. Structure and in Vivo Requirement of the Yeast Spt6 SH2 Domain. *Journal of Molecular Biology* 2009; **389** (1): 211–225.
- Dermody JL, Dreyfuss JM, Villén J, Ogundipe B, Gygi SP, Park PJ, Ponticelli AS et al. Unphosphorylated SR-Like Protein Npl3 Stimulates RNA Polymerase II Elongation. *PLoS ONE* 2008; **3** (9): e3273.
- Dichtl B, Blank D, Sadowski M, Hübner W, Weiser S, Keller W. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J.* 2002; **21** (15): 4125–4135.
- Dodson CA, Ferguson N, Rutherford TJ, Johnson CM, Fersht AR. Engineering a two-helix bundle protein for folding studies. *Protein Engineering Design and Selection* 2010; **23** (5): 357–364.
- Domínguez-Sánchez MS, Barroso S, Gómez-González B, Luna R, Aguilera A, Pearson CE. Genome Instability and Transcription Elongation Impairment in Human Cells Depleted of THO/TREX. *PLoS Genet* 2011; **7** (12): e1002386.
- Dufu K, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, Reed R. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes & Development* 2010; **24** (18): 2043–2053.
- Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol. Cell* 2012; **45** (1): 111–122.
- Fabrega C, Shen V, Shuman S, Lima CD. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Molecular Cell* 2003; **11** (6): 1549–1561.
- Fasken MB, Stewart M, Corbett AH. Functional Significance of the Interaction between the mRNA-binding Protein, Nab2, and the Nuclear Pore-associated Protein, Mlp1, in mRNA Export. *Journal of Biological Chemistry* 2008; **283** (40): 27130–27143.

- Faza MB, Kemmler S, Jimeno S, Gonzalez-Aguilera C, Aguilera A, Hurt E, Panse VG et al. Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery. *The Journal of Cell Biology* 2009; **184** (6): 833–846.
- Fong N, Öhman M, Bentley DL. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nat Struct Mol Biol* 2009; **16** (9): 916–922.
- Gallardo M. Nab2p and the Thp1p-Sac3p Complex Functionally Interact at the Interface between Transcription and mRNA Metabolism. *Journal of Biological Chemistry* 2003; **278** (26): 24225–24232.
- Gallardo M, Aguilera A. A new hyperrecombination mutation identifies a novel yeast gene, THP1, connecting transcription elongation with mitotic recombination. *Genetics* 2001; **157** (1): 79–89.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5** (10): R80.
- Ghosh A, Shuman S, Lima CD. Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. *Molecular Cell* 2011; **43** (2): 299–310.
- Gilbert W, Guthrie C. The Glc7p Nuclear Phosphatase Promotes mRNA Export by Facilitating Association of Mex67p with mRNA. *Molecular Cell* 2004b; **13** (2): 201–212.
- Gilbert W, Siebel CW, Guthrie C. Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA* 2001; **7** (2): 302–313.
- Goldstrohm AC, Albrecht TR, Sune C, Bedford MT, Garcia-Blanco MA. The Transcription Elongation Factor CA150 Interacts with RNA Polymerase II and the Pre-mRNA Splicing Factor SF1. *Molecular and Cellular Biology* 2001; **21** (22): 7617–7628.
- Gómez-González B, García-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marín A, Foiani M et al. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* 2011; **30** (15): 3106–3119.
- Gonzalez-Aguilera C, Tous C, Babiano R, La Cruz J de, Luna R, Aguilera A. Nab2 functions in the metabolism of RNA driven by polymerases II and III. *Molecular Biology of the Cell* 2011b; **22** (15): 2729–2740.
- Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A. The THP1-SAC3-SUS1-CDC31 Complex Works in Transcription Elongation-mRNA Export Preventing RNA-mediated Genome Instability. *Molecular Biology of the Cell* 2008; **19** (10): 4310–4318.
- Graveley BR, Hertel KJ. SR Proteins. In *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons 2005.
- Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, Stutz F et al. Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proceedings of the National Academy of Sciences* 2006; **103** (44): 16376–16381.
- Häcker S, Krebber H. Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J. Biol. Chem.* 2004; **279** (7): 5049–5052.
- Hackmann A, Gross T, Baierlein C, Krebber H. The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits. *EMBO Rep* 2011; **12** (10): 1024–1031.
- Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* 2004; **11** (5): 394–403.
- Hartzog GA, Fu J. The Spt4–Spt5 complex: A multi-faceted regulator of transcription elongation. *RNA polymerase II Transcript Elongation* 2013; **1829** (1): 105–115.

- Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM et al. Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J.* 2002; **21** (7): 1800–1810.
- Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R et al. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J* 2012; **31** (12): 2784–2797.
- Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, Stewart M, Dargemont C et al. Structural Requirements for the Ubiquitin-associated Domain of the mRNA Export Factor Mex67 to Bind Its Specific Targets, the Transcription Elongation THO Complex Component Hpr1 and Nucleoporin FXFG Repeats. *Journal of Biological Chemistry* 2009; **284** (26): 17575–17583.
- Hobeika M, Brockmann C, Iglesias N, Gwizdek C, Neuhaus D, Stutz F, Stewart M et al. Coordination of Hpr1 and Ubiquitin Binding by the UBA Domain of the mRNA Export Factor Mex67. *Molecular Biology of the Cell* 2007; **18** (7): 2561–2568.
- Hsin J, Sheth A, Manley JL. RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3' End Processing. *Science* 2011; **334** (6056): 683–686.
- Huertas P, Aguilera A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* 2003; **12** (3): 711–721.
- Hurt E, Luo M, Röther S, Reed R, Strässer K. Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proc. Natl. Acad. Sci. U.S.A.* 2004; **101** (7): 1858–1862.
- Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Dach E von, Corbett AH, Dargemont C et al. Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes & Development* 2010; **24** (17): 1927–1938.
- Jaehning JA. The Paf1 complex: Platform or player in RNA polymerase II transcription? *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2010; **1799** (5-6): 379–388.
- Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A et al. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* 2004; **21** (11): 947–962.
- Jimeno S, Luna R, Garcia-Rubio M, Aguilera A. Tho1, a Novel hnRNP, and Sub2 Provide Alternative Pathways for mRNP Biogenesis in Yeast THO Mutants. *Molecular and Cellular Biology* 2006; **26** (12): 4387–4398.
- Jimeno S, Rondón AG, Luna R, Aguilera A. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J.* 2002; **21** (13): 3526–3535.
- Johnson SA, Cubberley G, Bentley DL. Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3' End Processing Factor Pcf11. *Molecular Cell* 2009; **33** (2): 215–226.
- Johnson SA, Kim H, Erickson B, Bentley DL. The export factor Yra1 modulates mRNA 3' end processing. *Nat Struct Mol Biol* 2011; **18** (10): 1164–1171.
- Katahira J, Katahira J, Strässer K, Podtelejnikov A, Mann M, Jung JU, Hurt E. The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *EMBO J.* 1999; **18** (9): 2593–2609.
- Katahira J, Inoue H, Hurt E, Yoneda Y. Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. *EMBO J* 2009; **28** (5): 556–567.
- Katahira J, Sträßer K, Podtelejnikov A, Mann M, Jung JU, Hurt E, Katahira J et al. The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *EMBO J* 1999; **18** (9): 2593–2609.

- Kelly SM, Pabit SA, Kitchen CM, Guo P, Marfatia KA, Murphy TJ, Corbett AH et al. Recognition of polyadenosine RNA by zinc finger proteins. *Proceedings of the National Academy of Sciences* 2007; **104** (30): 12306–12311.
- Kessler SH, Sachs AB. RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Mol. Cell. Biol.* 1998; **18** (1): 51–57.
- Kim M, Suh H, Cho E, Buratowski S. Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. *Journal of Biological Chemistry* 2009; **284** (39): 26421–26426.
- Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences* 2011; **108** (2): 546–550.
- Knop M, Finger A, Braun T, Hellmuth K, Wolf DH. Der1, a novel protein specifically required for endoplasmic reticulum degradation in yeast. *EMBO J.* 1996; **15** (4): 753–763.
- Köhler A, Hurt E. Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* 2007; **8** (10): 761–773.
- Kress TL, Krogan NJ, Guthrie C. A Single SR-like Protein, Npl3, Promotes Pre-mRNA Splicing in Budding Yeast. *Molecular Cell* 2008; **32** (5): 727–734.
- Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA Polymerase II CTD Phosphatase. *Molecular Cell* 2004; **14** (3): 387–394.
- Krogan NJ, Peng W, Cagney G, Robinson MD, Haw R, Zhong G, Guo X et al. High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* 2004; **13** (2): 225–239.
- Kuhlmann SI, Valkov E, Stewart M. Structural basis for the molecular recognition of polyadenosine RNA by Nab2 Zn fingers. *Nucleic Acids Research* 2013.
- Kyburz A, Sadowski M, Dichtl B, Keller W. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic Acids Res.* 2003; **31** (14): 3936–3945.
- Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; **227** (5259): 680–685.
- Leaw CL, Ren EC, Choong ML. Hcc-1 is a novel component of the nuclear matrix with growth inhibitory function. *CMLS, Cell. Mol. Life Sci.* 2004; **61** (17): 2264–2273.
- Lee DC, Aitchison JD. Kap104p-mediated nuclear import. Nuclear localization signals in mRNA-binding proteins and the role of Ran and Rna. *J. Biol. Chem.* 1999; **274** (41): 29031–29037.
- Lei EP, Krebber H, Silver PA. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev.* 2001; **15** (14): 1771–1782.
- Lei EP, Stern CA, Fahrenkrog B, Krebber H, Moy TI, Aebi U, Silver PA et al. Sac3 is an mRNA export factor that localizes to cytoplasmic fibrils of nuclear pore complex. *Mol. Biol. Cell* 2003; **14** (3): 836–847.
- Lei H, Dias AP, Reed R. Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex. *Proceedings of the National Academy of Sciences* 2011; **108** (44): 17985–17990.
- Lei H, Zhai B, Yin S, Gygi S, Reed R. Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export. *Nucleic Acids Research* 2013; **41** (4): 2517–2525.
- Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA et al. The Specificity and Topology of Chromatin Interaction Pathways in Yeast. *Molecular Cell* 2011; **42** (4): 536–549.

- Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL et al. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol. Cell* 2002; **9** (5): 1101–1111.
- Lidschreiber M, Leike K, Cramer P. Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. *Molecular and Cellular Biology* 2013; **33** (19): 3805–3816.
- Linder P, Jankowsky E. From unwinding to clamping — the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 2011; **12** (8): 505–516.
- Lindstrom DL, Squazzo SL, Muster N, Burckin TA, Wachter KC, Emigh CA, McCleery JA et al. Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. *Molecular and Cellular Biology* 2003; **23** (4): 1368–1378.
- Liu Q, Dreyfuss G. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* 1995; **15** (5): 2800–2808.
- Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, Ranish J et al. Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. *Molecular and Cellular Biology* 2009; **29** (17): 4852–4863.
- Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods* 2001; **25** (4): 402–408.
- Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H et al. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 2010; **17** (10): 1195–1201.
- Lund MK, Guthrie C. The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol. Cell* 2005; **20** (4): 645–651.
- Mackellar AL, Greenleaf AL. Cotranscriptional Association of mRNA Export Factor Yra1 with C-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry* 2011; **286** (42): 36385–36395.
- Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.* 2008; **65** (7-8): 1099–1122.
- Margaritis T, Holstege FC. Poised RNA Polymerase II Gives Pause for Thought. *Cell* 2008; **133** (4): 581–584.
- Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P. Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *EMBO J* 2011; **30** (7): 1302–1310.
- Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.* 2005; **19** (13): 1512–1517.
- Ma WK, Cloutier SC, Tran EJ. The DEAD-box Protein Dbp2 Functions with the RNA-Binding Protein Yra1 to Promote mRNP Assembly. *Journal of Molecular Biology* 2013; **425** (20): 3824–3838.
- Mayer A, Heidemann M, Lidschreiber M, Schrieck A, Sun M, Hintermair C, Kremmer E et al. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 2012a; **336** (6089): 1723–1725.
- Mayer A, Lidschreiber M, Siebert M, Leike K, Söding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* 2010; **17** (10): 1272–1278.
- Mayer A, Schrieck A, Lidschreiber M, Leike K, Martin DE, Cramer P. The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol. Cell. Biol.* 2012b; **32** (7): 1321–1331.
- Mayr C, Bartel DP. Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cellst. *Cell* 2009; **138** (4): 673–684.

- McBride AE, Cook JT, Stemmler EA, Rutledge KL, McGrath KA, Rubens JA. Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its function, nuclear export, and intranuclear protein interactions. *J. Biol. Chem.* 2005; **280** (35): 30888–30898.
- Meinhart A, Kamenski T, Hoepfner S, Baumli S, Cramer P. A structural perspective of CTD function. *Genes Dev.* 2005; **19** (12): 1401–1415.
- Millhouse S, Manley JL. The C-Terminal Domain of RNA Polymerase II Functions as a Phosphorylation-Dependent Splicing Activator in a Heterologous Protein. *Molecular and Cellular Biology* 2005; **25** (2): 533–544.
- Morris DPGA. The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry* 2000; **275** (51): 39935–39943.
- Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL et al. Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. *Molecular Cell* 2009; **34** (2): 168–178.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M et al. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* 2008a; **320** (5881): 1344–1349.
- Nissan TA, Galani K, Maco B, Tollervey D, Aebi U, Hurt E. A pre-ribosome with a tadpole-like structure functions in ATP-dependent maturation of 60S subunits. *Mol. Cell* 2004; **15** (2): 295–301.
- Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA et al. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* 2005; **12** (2): 144–151.
- Noble KN, Tran EJ, Alcazar-Roman AR, Hodge CA, Cole CN, Wentz SR. The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes & Development* 2011; **25** (10): 1065–1077.
- Nojima T, Hirose T, Kimura H, Hagiwara M. The Interaction between Cap-binding Complex and RNA Export Factor Is Required for Intronless mRNA Export. *Journal of Biological Chemistry* 2007; **282** (21): 15645–15651.
- Oeffinger M, Zenklusen D. To the pore and through the pore: A story of mRNA export kinetics. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2012; **1819** (6): 494–506.
- O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ. Comparison of sample preparation methods for CHIP-chip assays. *BioTechniques* 2006; **41** (5): 577–580.
- Olins DE, Olins AL. Chromatin history: our view from the bridge. *Nat. Rev. Mol. Cell Biol.* 2003; **4** (10): 809–814.
- Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B et al. Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. *Cell* 2010; **143** (6): 1018–1029.
- Peña Á, Gewartowski K, Mroczek S, Cuéllar J, Szykowska A, Prokop A, Czarnocki-Cieciura M et al. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO J* 2012; **31** (6): 1605–1616.
- Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes & Development* 2011; **25** (17): 1770–1782.
- Qiu H, Hu C, Gaur NA, Hinnebusch AG. Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J* 2012a; **31** (16): 3494–3505.
- Qiu H, Hu C, Hinnebusch AG. Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. *Molecular Cell* 2009; **33** (6): 752–762.

- Rando OJ, Ahmad K. Rules and regulation in the primary structure of chromatin. *Curr. Opin. Cell Biol.* 2007; **19** (3): 250–256.
- RDevelopmentCoreTeam. R: a language and environment for statistical computing, Vienna Austria 2009.
- Reed R, Hurt E. A Conserved mRNA Export Machinery Coupled to pre-mRNA Splicing. *Cell* 2002; **108** (4): 523–531.
- Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes & Development* 2009; **23** (11): 1247–1269.
- Rodríguez-Navarro S, Fischer T, Luo M, Antúnez O, Brettschneider S, Lechner J, Pérez-Ortín JE et al. Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. *Cell* 2004; **116** (1): 75–86.
- Rodríguez-Navarro S, Hurt E. Linking gene regulation to mRNA production and export. *Current Opinion in Cell Biology* 2011; **23** (3): 302–309.
- Rondón AG, Jimeno S, Aguilera A. The interface between transcription and mRNP export: From THO to THSC/TREX-2. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2010; **1799** (8): 533–538.
- Rother S, Burkert C, Brunger KM, Mayer A, Kieser A, Strasser K. Nucleocytoplasmic shuttling of the La motif-containing protein Sro9 might link its nuclear and cytoplasmic functions. *RNA* 2010; **16** (7): 1393–1401.
- Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J et al. THO/Sub2p Functions to Coordinate 3'-End Processing with Gene-Nuclear Pore Association. *Cell* 2008; **135** (2): 308–321.
- Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* 2000; **148** (4): 635–651.
- Saguez C, Schmid M, Olesen JR, Ghazy MAE, Qu X, Poulsen MB, Nasser T et al. Nuclear mRNA Surveillance in THO/sub2 Mutants Is Triggered by Inefficient Polyadenylation. *Molecular Cell* 2008; **31** (1): 91–103.
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science* 2008; **320** (5883): 1643–1647.
- Santos-Rosa H, Moreno H, Simos G, Segref A, Fahrenkrog B, Panté N, Hurt E et al. Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores. *Mol. Cell. Biol.* 1998; **18** (11): 6826–6838.
- Senger B, Simos G, Bischoff FR, Podtelejnikov A, Mann M, Hurt E. Mtr10p functions as a nuclear import receptor for the mRNA-binding protein Npl3p. *EMBO J.* 1998; **17** (8): 2196–2207.
- Shen Z, St-Denis A, Chartrand P. Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4-Spt5/DSIF promotes mRNA localization to the yeast bud. *Genes & Development* 2010; **24** (17): 1914–1926.
- Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 1989; **122** (1): 19–27.
- Strasser K, Bassler J, Hurt E. Binding of the Mex67p/Mtr2p Heterodimer to Fxfg, Glfg, and Fg Repeat Nucleoporins Is Essential for Nuclear mRNA Export. *The Journal of Cell Biology* 2000; **150** (4): 695–706.

- Strässer K, Hurt E. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J.* 2000a; **19** (3): 410–420.
- Strässer K, Hurt E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* 2001; **413** (6856): 648–652.
- Strässer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodríguez-Navarro S, Rondón AG et al. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 2002; **417** (6886): 304–308.
- Sugiura T, Sakurai K, Nagano Y. Intracellular characterization of DDX39, a novel growth-associated RNA helicase. *Exp. Cell Res.* 2007; **313** (4): 782–790.
- Tian B, Manley JL. Alternative cleavage and polyadenylation: the long and short of it. *Trends in Biochemical Sciences* 2013; **38** (6): 312–320.
- Tietjen JR, Zhang DW, Rodríguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X et al. Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 2010; **17** (99): 1154–1161.
- Toedling J, Sklyar O, Huber W. Ringo – an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics* 2007; **8** (1): 221.
- Topisirovic I, Svitkin YV, Sonenberg N, Shatkin AJ. Cap and cap-binding proteins in the control of gene expression. *WIREs RNA* 2011; **2** (2): 277–298.
- Tran EJ, Zhou Y, Corbett AH, Wentz SR. The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. *Mol. Cell* 2007; **28** (5): 850–859.
- Tuck AC, Tollervey D. A Transcriptome-wide Atlas of RNP Composition Reveals Diverse Classes of mRNAs and lncRNAs. *Cell* 2013; **154** (5): 996–1009.
- van Werven FJ, Timmers HTM. The use of biotin tagging in *Saccharomyces cerevisiae* improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Res.* 2006; **34** (4): e33.
- Vinciguerra P, Iglesias N, Camblong J, Zenklusen D, Stutz F. Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. *EMBO J* 2005; **24** (4): 813–823.
- Vojnic E, Mourão A, Seizl M, Simon B, Wenzek L, Larivière L, Baumli S et al. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* 2011; **18** (4): 404–409.
- Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Micro* 2011; **9** (2): 85–98.
- West ML, Corden JL. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 1995; **140** (4): 1223–1233.
- Wilkening S, Pelechano V, Jarvelin AI, Tekkedil MM, Anders S, Benes V, Steinmetz LM et al. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Research* 2013; **41** (5): e65.
- Will CL, Luhrmann R. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology* 2011; **3** (7): a003707.
- Windgassen M, Krebber H. Identification of Gbp2 as a novel poly(A)⁺ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. *EMBO Rep* 2003; **4** (3): 278–283.
- Windgassen M, Sturm D, Cajigas IJ, Gonzalez CI, Seedorf M, Bastians H, Krebber H et al. Yeast Shuttling SR Proteins Npl3p, Gbp2p, and Hrb1p Are Part of the Translating mRNPs, and Npl3p Can Function as a Translational Repressor. *Molecular and Cellular Biology* 2004; **24** (23): 10479–10491.

- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature* 2009; **457** (7232): 1033–1037.
- Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. *Molecular Cell* 2006; **21** (2): 227–237.
- Yao W, Roser D, Köhler A, Bradatsch B, Bassler J, Hurt E. Nuclear export of ribosomal 60S subunits by the general mRNA export receptor Mex67-Mtr2. *Mol. Cell* 2007; **26** (1): 51–62.
- Yun CY, Fu X. Conserved Sr Protein Kinase Functions in Nuclear Import and Its Action Is Counteracted by Arginine Methylation in *Saccharomyces cerevisiae*. *The Journal of Cell Biology* 2000; **150** (4): 707–718.
- Zacher B, Kuan P, Tresch A. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. *BMC Bioinformatics* 2010; **11** (1): 194.
- Zhang DW, Rodríguez-Molina JB, Tietjen JR, Nemec CM, Ansari AZ. Emerging Views on the CTD Code. *Genetics Research International* 2012; **2012** (9): 1–19.
- Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* 1999; **63** (2): 405–445.
- Zhou K, Kuo WHW, Fillingham J, Greenblatt JF. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proceedings of the National Academy of Sciences* 2009; **106** (17): 6956–6961.
- Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. *Nature* 2002; **419** (6903): 182–185.